Develop New or Improved Instruments and Technologies for Use in Research and Medicine

IN THIS SECTION:

SCIENCE ADVANCES

SCIENCE CAPSULES (page 276)
· Flourescent Brain Cells Help Track Master Reproductive Hormone
· Detailed Images Achieved with New Magnetic Resonance Imaging (MRI) Technique
· Improved Prospects for Gene Therapy Requiring Large Genes
· New Method to More Quickly Study Cellular Regulatory Processes
· Automatable Method for Protein Identification
· Efficient Production of Dopamine Nerve Cells
· Hearing Aids
· Altruism Motivates Participation in Schizophrenia Research
· Assuring Safe Feeding Tube Placement
· A Database of SNPs Will Shed Light on Human Variation and Disease
· A New Method for Searching for Patterns in Proteins
· Complete Genome Data for More Than 700 Organisms Facilitates Comparative Analyses
· Next Generation Internet Implementation to Serve Visible Human Datasets
· Remote Treatment Planning System
· Mammography for the Next Generation Internet
· Multilateral Initiative on Malaria
· Tribal Connections
· Internet Connectivity Performance Evaluation
· New Imaging Instrumentation for Diagnosis of Eye Diseases
· Intranasal Administration of Naked TGF-B DNA May Be Used to Treat Inflammatory Diseases
· Early Detection of Cancer Cells
· Novel SPECT Ligands to Image the Serotonin Transporter in Human Brain
· New Technique Eliminates a Step in Making Genetically-Altered Mice
· RNA Interference: A New Method for Controlling Gene Expression

STORIES OF DISCOVERY (page 286)
· MEDLINE and MEDLINEplus: A Continuing Story of Discovery
· Genome-scale Analyses: cDNA and Tissue Chips
· Single Nucleotide Polymorphisms (SNPs): New Tools for Tracing Inherited Diseases

The New Immigrant Study:
Understanding the Background, Skills, and Impact of Immigration on U.S. Society

Background: Despite the importance of immigration to the past, present, and future well-being of the Nation, no survey has ever contained a nationally representative sample of new legal immigrants and been designed to follow their progress over time. This is largely due to the difficulties and costs involved: conducting such a survey entails significant challenges in finding, interviewing, and retaining respondents who are apt to be highly mobile within the U.S. and abroad. To help researchers address a range of issues likely to affect the health of mothers, children and families for decades to come, the NIH was joined by, among others, the Immigration and Naturalization Service and the National Science Foundation in supporting the development of the New Immigrant Survey (NIS). The NIS examines the characteristics of legal permanent residents (e.g., green card holders) in the United States, and is the first survey to target questions specifically to the "green card" population concerning health, schooling, language skills, earnings, past experience in illegal status, and economic gains from immigrating.

Advance: As an important step in developing the full-scale NIS, the investigators conducted a highly successful pilot study. In terms of methodology, the NIS pilot study (NIS-P) showed that the complex procedures for locating, interviewing, and reinterviewing the target population are sound, and that the quality of the data obtained is high. In terms of content, the study demonstrated that the previous information about non-citizens has left misleading impressions concerning legal immigrants in the US. The data also revealed some important new facts. For example, when comparing individuals with postgraduate education, adult green card holders are three times more likely to have schooling beyond college than U.S. natives. On the other hand, when comparing individuals with less education, green card holders are also about three times more likely to have completed fewer than nine years of schooling than U.S. natives. Most new legal immigrants are not new to the United States: two-thirds have had previous experience living here, and at least one in five entered illegally at some point in the past. When compared to their last job abroad, about a quarter of the new legal immigrants experienced an initial drop in earnings. Nevertheless, immigrants overall experienced large increases in their annual earnings – an average of $10,000 for men and $6,000 for women. Average earnings of employed new immigrant men fall short of those enjoyed by U.S.-born men, but employed new-immigrant women earn *more* than their U.S.-born counterparts.

Implications: The NIS-P has proven that it is possible to overcome methodological and logistical difficulties in developing a longitudinal study of U.S. immigrants, and promises that the full-scale survey will revolutionize immigrant research in the U.S. The study also provides the first clear insights into the characteristics of legal immigrants to this country. The data serve as a critically important resource for policy-makers and for scientists analyzing the causes and consequences of immigration. The data also contribute significantly to our fundamental understanding of how immigration fuels population growth, economic productivity, the rapidly increasing diversity of our population, and the health and well-being of immigrant populations.

Jasso G, Massey DS, Rosenzwieg MR, Smith JP: The new immigrant survey pilot (NIS-P): overview and new findings about US legal immigrants at admission. Demography, 37(1):127-38. 2000.

The Mouse Intracytoplasmic Sperm Injection Model: No Adverse Effects of Bypassing Conventional Sperm Formation and Fertilization

Background: Intracytoplasmic sperm injection (ICSI) has become the method of choice to help overcome male infertility and is widely used by infertility clinics around the world. The ICSI method allows clinicians to inject either mature or immature sperm directly into an oocyte, or egg cell. This technique bypasses many natural biological processes involved in fertilization and allows sperm from an infertile male, which could not enter into an egg under normal or *in vitro* conditions, to be simply injected into the egg's cytoplasm to initiate embryo development. After transfer to a uterus, the embryo can develop into live offspring. Researchers are particularly interested if repeated use of the ICSI method, especially when immature sperm are used, will have any harmful long-term effects. Because it takes nearly 20 years to produce at least one human generation, researchers use mice to quickly determine if repeated use of the procedure will have serious consequences on the fertility and behavior of offspring.

Advance: A new study shows that immature sperm can be injected directly into mouse eggs to produce live offspring. This ICSI procedure has now been carried forward through five generations of such live offspring, using immature sperm cells from each generation of mice. In this process, the scientists found the ICSI mice had no apparent differences in their growth and fertility when compared to controls born after natural mating. In addition, the researchers found that the subsequent generations of ICSI mice had no clear deficits in terms of elementary reasoning, emotionality, spatial learning and memory.

Implications: ICSI methods are used annually in tens of thousands of cases worldwide to overcome infertility. It is reassuring to find that bypassing the natural steps in the fertilization process, including the full maturation of sperm had no cumulative or damaging effect upon the growth, fertility or behavior of subsequent generations of mice. It is important to note, however, that *normal fertile* mice were used in this study. While these studies are important, the apparent safety of ICSI in mice cannot be taken to mean that the procedure is completely harmless to humans: longitudinal studies are needed to answer this question.

Tamashiro KLK, Kimura Y, Blanchard RJ, Blanchard DC, Yanagimachi R: Bypassing spermiogenesis for several generations does not have detrimental consequences on the fertility and neurobehavior of offspring: a study using the mouse. Journal of Assisted Reproduction and Genetics, 16(6):315-24. 1999.

Robots Eavesdrop on Cellular Discussions

Background: Imagine visiting a library full of books that you couldn't read. In a sense, this is the scientific dilemma facing biologists across the globe. Researchers have in hand boatloads of genetic information – billions of DNA letters that spell out the instructions for life in organisms as diverse as yeast, worms, flies, and humans. The problem is that, to a great degree, no one knows what all these genes do. And even in the cases where scientists do know, even more puzzling is how cell parts communicate with each other, often through physical contact. While scientists have developed powerful approaches to determine which of the thousands of genes are "turned on" in a particular cell, they haven't had a "guidebook" to tell them which gene products interact physically.

Advance: For the first time, researchers working with the model organism *Saccharomyces cerevisiae* (baker's yeast) have figured out a way to record the "conversations" taking place simultaneously between thousands of molecules inside a single cell. Using robots to monitor the goings-on of thousands of individual yeast cells growing on a small plastic grid, these scientists have accomplished a biological milestone in determining which molecules in a cell "talk" to others by making physical contact. To achieve this feat, the researchers made use of robotic devices to automate state-of-the-art, but commonly used, molecular biological techniques.

Implications: "Listening in" on which proteins physically talk to other proteins is a critical task for researchers since all cells rely on extensive and ongoing molecular discussions to carry out life's functions – everything from breathing to memory. When the complete, ordered sequence of the human genome is available to researchers in the next couple of years, a similar strategy will likely be possible using human cells. In the near term, scientists all over the world studying yeast cells as a model for understanding human health and disease will be able to use this information to advance their research.

Uetz P, Giot L, Cagney G, et al: A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. Nature, 403(6770):623-7. 2000.

Targeted DNA Insertion May Aid Gene Therapy

Background: For decades, scientists have tried to insert genes into precise locations within the genetic material of laboratory test organisms. Such experiments help them better understand the function of the inserted genes and enable them to disrupt other genes. Gene therapy also relies on the insertion of corrective genes into the cells of living humans. With most current methods, the genes insert randomly, so scientists cannot control whether the genes will function in their new location. Such random insertions can even lead to potentially disease-causing mutations.

Advance: New research indicates that it is possible to insert genes into any desired location in the genetic material of a host organism. Geneticists accomplished this by harnessing the targeting mechanism of a portion of DNA called an intron. Often called "junk DNA," introns normally do not code for proteins, but exist in the middle of genes that do. Some introns can recognize a sequence of DNA and insert themselves precisely into it. They do this by matching up a region of their own genetic sequence with the targeted sequence. To see whether it is possible to construct introns to target any desired gene, researchers chose two clinically relevant sequences to target: one in a gene in HIV, the virus that causes AIDS, and another in a human gene called CCR5 that encodes a protein necessary for HIV infection. People who have mutations in their CCR5 genes are resistant to HIV infection, leading scientists to believe that disrupting the CCR5 gene may be an effective anti-AIDS therapy. The scientists modified a type of intron found in bacteria to generate a variety of introns with different genetic sequences in their targeting regions. They found that more than a dozen of these introns inserted themselves into the intended positions in the target genes. The research was done using human cells in a laboratory, but the scientists ultimately hope to use the technique to treat HIV infection in people.

Implications: If further tests are successful, the method could be used both to disrupt specific genes and to add new genes at desired sites. This could advance all sorts of biomedical research, ranging from studies on basic gene function to the development of antiviral and antibacterial drugs. The work may also enhance the delivery of genes for gene therapy. Currently, the specialized viruses usually used to deliver corrective genes insert their cargo at random sites in human DNA. If the new method could deliver therapeutic genes to specified sites, it could make gene therapy safer and more effective. The work is an excellent example of an unexpected outcome from basic biomedical research. While researchers were carrying out basic research on these itrons and how they related to gene structure, they reaped the benefit if their use in practical applications as well.

Guo H, Karberg M, Long M, Jones JP III, Sullenger B, Lambowitz AM: Group II introns designed to insert into therapeutically relevant DNA target sites in human cells. Science, 289(5478):452-7. 2000.

Strauss E: Targeting intron insertion into DNA. Science, 289(5478):374. 2000.

Simple Breath Test Predicts Gene-Linked Drug Response

Background: Pharmacogenetics is a blossoming area of research aimed at connecting a person's genetic make-up with his or her response to medicines. As drugs move through the body, they interact with thousands of molecules called proteins, each encoded by a different gene. Because each person is genetically unique, tiny differences in these proteins can affect how medicines do their jobs in the body, since some of these proteins work to get rid of medicines, while others help medicines do their jobs. For example, certain drugs used to treat cancer can have widely variable effects in patients, and many of these treatments have serious toxicities. On occasion, patients are literally poisoned because their bodies cannot get rid of, or "clear," cancer treatments in a timely manner. Patients given an identical dose of the commonly used chemotherapy drug Taxotere? (docetaxel), for instance, display wide variation in the amount of time it takes their bodies to clear this medicine. Scientists have suspected that gene differences may account for at least some of the variability in response to docetaxel.

Advance: Now, pharmacogenetics researchers have confirmed this suspicion using a simple "breath test" that measures an individual patient's ability to break down docetaxel. Several years ago, the same basic researchers came up with the idea for the breath test, which measures the strength of a drug-metabolizing protein nicknamed "CYP3A4." The CYP3A4 protein chews up many different drugs, including the antibiotic erythromycin. Researchers use the breath test to gauge patient's CYP3A4 activity by injecting a tiny amount of erythromycin containing trace levels of the radioisotope ^{14}C, and then measuring the amount of radioactive carbon dioxide the patient exhales 20 minutes later. (The very small amount of radioactivity poses no danger to the patient or health care provider.) In a small clinical study, the researchers successfully used the breath test to track patients' breakdown of docetaxel. Underscoring the predictive value of the simple test, the scientists found that the patients with the lowest scores on the breath test were the ones who suffered the greatest docetaxel toxicity.

Implications: The new work introduces an important potential role for the breath test in predicting toxicity caused by a widely used cancer drug. Since many medicines are metabolized by the CYP3A4 protein, the breath test may prove useful in the clinic as a rapid and easy way to predict individual patients' responses to other drugs. Since previous blood tests failed to predict docetaxel toxicity ahead of time, the breath test may offer a promising tool to help physicians administer this drug more safely. Makers of the breath test are currently seeking FDA approval – a step that will mark another practical use of pharmacogenetics in the clinic.

Hirth J, Watkins PB, Strawderman M, Schott A, Bruno R, Baker LH: The effect of an individual's cytochrome CYP3A4 activity on docetaxel clearance. Clinical Cancer Research, 6:1255-8. 2000.

Laboratory-Grown Heart Valves Show Potential as Valve Replacements

Background: Heart valves are flap-like structures that help regulate blood flow through the heart. Malfunctioning valves can seriously impede heart function. More than 60,000 patients in the United States receive heart valve replacements each year. Although their overall performance is excellent, major problems are still associated with each type of valve replacement presently used. Tissue-engineering – a new technology that applies the principles of engineering to create functional tissues and organs that can replace damaged or diseased natural tissues and organs – has potential for producing less problematic replacement heart valves. Until now, heart valves generated by tissue-engineering methods did not perform well once they were implanted.

Advance: Recently, a group of researchers, using a new tissue-engineering technique, "grew" heart valves in their laboratory and implanted them into six lambs. The valves functioned satisfactorily for up to five months. Even more important, the engineered valves gradually evolved to resemble natural valves in terms of several mechanical and structural characteristics. The valves were grown from bioabsorbable materials seeded with cells from blood vessels of the lambs to receive the implant. Using a special laboratory system they developed, the researchers "conditioned" the valves by growing them in an environment that simulated the pressure and flow of blood in the body. The new technique produced more mature valves with greater ability to function than valves previously generated.

Implications: Despite four decades of research and development, an ideal heart valve replacement has not yet been developed. Further development of tissue-engineering systems like the one developed and used in this study could result in heart valves that are far better than the ones in use today. Tissue-engineered heart valves might be able to function for the remainder of a patient's life, provide ongoing tissue alteration and repair as needed, and, in the case of pediatric patients, grow as the patients grow.

Hoerstrup SP, Sodian R, Daebritz S, et al: Functional living trileaflet heart valves grown in vitro. Circulation, 102(19)Supplement III:44-9. 2000.

Clinical Study Shows that New Treatment May Improve Feasibility of Bone Marrow Transplantation

Background: Bone marrow transplantation is frequently used in other countries to treat patients with sickle cell disease and thalassemia. This procedure is rarely performed in such patients in the U.S. because the procedure carries the risk of painful and potentially fatal complications such as graft versus host disease (GVHD). Preparing patients for transplantation is also risky; it entails irradiating patients to destroy their bone marrow, thereby eradicating not only their disease but also their immune response.

Advance: In a search for new and safer ways to treat patients with blood diseases that have the potential to respond to bone marrow transplantation, scientists developed a treatment regimen consisting of less radiation than is used currently coupled with immunosuppressive drugs following transplantation. Because all of a recipient's blood cells do not have to be destroyed before new donor blood cells can be incorporated, procedures can be used that cause less damage to a patient's cells and therefore are less likely to be fatal than more destructive techniques.

The treatment was successful in dogs where transplanted cells were grafted successfully in 12 of 13 animals and none of the dogs developed GVHD. Preliminary results from 26 patients indicate that the procedure is well tolerated in humans and has the potential to be performed entirely in an outpatient setting.

Implications: Because the likelihood of GVHD increases with age and because the doses of ionizing radiation that traditionally are used can be toxic to the digestive system, bone marrow transplantation currently is limited to younger patients with good organ function. However, the new treatment protocol allows the patient population to be expanded. The current study is being performed using related recipient-donor pairs. Evaluation with unrelated donors is needed before the procedure will increase dramatically the number of patients who are eligible to receive bone marrow transplantation. Ultimately, the performance of the less damaging procedure in an outpatient setting also may result in major cost savings.

Sandmaier BM, McSweeny P, Yu C, Storb R: Nonmyeloablative transplants: preclinical and clinical results. Seminars in Oncology 27(2):78-81. 2000.

Storb R, Yu C, Zaucha M, et al: Stable mixed heamtopoietic chimerism in dogs given donor antigen, CTLA4Ig, and 100 cGy total body irradiation before and pharmacologic immunosuppression after marrow transplant. Blood, 94(7):2523-9. 1999.

Development of a Novel Technique To Analyze Membrane Proteins

Background: In an era when scientists are rapidly identifying new genes and the proteins that they code for, it has become more important than ever to quickly characterize the structure and function of these new proteins. One important class of proteins has remained resistant to characterization; those that are found in the lipid membranes that both surround the cell and form the barriers between its organelles or "little organs." Membrane-bound proteins, about 10 percent of all proteins in the cell, have an enormous variety of crucial functions. They can be receptors that detect hormones and other important signals, and act as channels that allow nutrients, ions, and waste products to cross in and out of the cell, and between organelles. Unfortunately, the well-established methods used to find the structure of most proteins, x-ray crystallography and nuclear magnetic resonance spectroscopy, have not been successful for these membrane-bound proteins.

Advance: Investigators have developed a widely-applicable technique that permits the rapid determination of the molecular mass of a full-length protein with high precision. Investigators have demonstrated the broad usefulness of this technique by employing it to perform mass spectrometric analysis of four entirely different membrane proteins from three bacterial organisms. Because of the precision in determining molecular mass, this technique can identify new proteins, detect even small errors in protein amino acid sequence of known proteins, and characterize important chemical modifications such as glycosylation or oxidation. Importantly, it can be used to identify new proteins in a crude tissue preparation without extensive purification, a step that has often proved difficult and time-consuming.

Implications: The completion of the human genome sequencing project, as well as the continued sequencing of a number of other genomes, will lead to the identification of a number of proteins of unknown structure and function. Rapid chromatographic separation of crude protein extracts with subsequent spectrometric analysis using this new technique can provide information necessary for determining protein structure and function. This will be especially useful in those cases, such as membrane proteins, where classical methods have failed.

le Coutre J, Whitelegge JP, Gross A, et al: Proteomics on full-length membrane proteins using mass spectrometry. Biochemistry, 39(15):4237-42. 2000.

A Useful Technique for Understanding Signaling in the Cell

Background: RNA-mediated genetic interference (RNAi) is a process in which targeted inhibition of gene expression can be mediated through double stranded RNA. Discovery of this phenomenon came from reverse genetic studies in *C. elegans,* a small roundworm. Recently, RNAi has been reported to work in fruitfly *Drosophila* as well as other invertebrates and plants. In contrast to *C. elegans*, there are a number of established cell lines from the fruitfly *Drosophila* offering many experimental advantages for biochemical studies. In addition, the *Drosophila* research community, both through the efforts of "genome projects" and of individual investigators, has accumulated a vast amount of data on the genetic and molecular organization of the *Drosophila* genome, as well as on the structure, expression and function of individual genes. Investigators are now using this approach to study signaling cascades by dissecting the well-characterized insulin signal transduction pathway.

Advance: Researchers have demonstrated that RNAi is capable of blocking the production of specific proteins using cell cultures from the fruitfly *Drosophila.* To date, researchers have used RNAi successfully to inhibit the production of or "knock-out" fifteen proteins, including kinases, phosphatases, adaptor proteins, and receptors. Data suggest that this approach is effective in inhibiting a broad range of proteins. RNAi was able to block the expression of two proteins already under study within the laboratory. One was an analog of a human protein called "nexin," one of a set of five related proteins that serves to sort out and direct intracellular movements of proteins. The other was an analog of common mammalian enzymes called kinases, a protein tyrosine kinase. Using RNAi, researchers were then able to study the relationship of one protein to the other. They established that the second protein functions as a kinase and precedes the site of activation, or phosphorylation, of the first protein. The investigators next verified that the RNAi could be used to dissect out signal transduction pathways by testing it in the well-characterized insulin signaling pathway. Inhibiting the expression of four distinct proteins in the insulin signaling pathway produced a predictable resulting response, confirming known steps in the pathway.

Implications: Intracellular signaling processes are fundamental mechanisms underlying health and disease. Here, researchers have demonstrated the applicability and efficacy of double-stranded RNA-mediated interference of gene expression in blocking the production of targeted proteins in the fruitfly. This method is technically simple, highly reproducible and quick results may be obtained within a few days. These results highlight the usefulness of RNAi in analyzing complex biochemical pathways and provide an effective method for determining the function of genes identified from the *Drosophila* genome sequencing project. These experiments may also serve as a stepping stone for adaptation of this technique to mammalian systems.

Clemens JC, Worby CA, Simonson-Leff N, et al: Use of double-stranded RNA interference in *Drosophila* cell lines to dissect signal transduction pathways. Proceedings of the National Academy of Sciences, 97(12):6499-6503. 2000.

Genome-Wide Screening of Protein Interactions

Background: The recent deciphering of the human genome will not have medical significance unless scientists can determine how these genes function in the human body. The genes and their protein products orchestrate the complex activities within each cell. Understanding this intricate web of interlocking protein interactions requires new approaches to attack the problem on a larger scale.

Advance: In collaboration with a biotechnology company, researchers at the University of Washington have developed an automated system for genome-wide screening of thousands of protein interactions that occur within a cell. The scientists tested this new technology on the single-celled organism *Saccharomyces cerevisiae,* commonly known as baker's yeast, which serves as a model organism for understanding human biology and disease. Because transient pairings between proteins lie at the heart of most cellular activities, researchers can obtain clues to the function of a poorly understood protein by determining if it interacts with another protein of known function. Yeast geneticists had previously developed a technique, known as the two-hybrid assay, to identify single protein-protein interactions. Now, using automated technologies, scientists have been able to apply the two-hybrid assay to screen the entire yeast genome and to simultaneously assess hundreds of protein interactions in yeast cells and determine how they function. Using high-throughput technology and a database system that identifies novel protein interactions, the researchers detected nearly 1,000 putative protein-protein associations, many of which involved poorly understood yeast proteins that can now be studied in new contexts.

Implications: The completion of the first protein interaction map for an entire model organism – yeast – is a landmark achievement. It is the first time that the protein interactions of a complete genome have been identified and mapped. This type of high-throughput protein-interaction screening can now be applied to the genomes of additional organisms. By discovering the functions of genes and proteins in simple organisms such as yeast, scientists can gain a clearer picture of how similar genes function in humans. This information will be valuable in determining the correlation between specific genes and complex diseases such as metabolic disorders, cancer, and autoimmune diseases.

Uetz P, Giot L, Cagney G, et al: A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. Nature, 403(6770):623-7. 2000.

Fingerprinting Bacteria

Background: Identification of pathogenic bacteria is important for medical diagnosis, food and water safety, anti-bioterrorism, and basic biomedical research. Among the approximately 4,000 bacterial species identified to date, roughly 200 are pathogenic to humans and cause diseases that range in severity from moderate discomfort (in cases of mild food poisoning) to life-threatening. Scientists estimate that millions of additional bacterial species have not yet been discovered. Methods for rapidly identifying bacteria would allow physicians to administer early and appropriate therapies to their patients.

Advance: Scientists at the NIH Flow Cytometry Resource Center, Los Alamos National Laboratory, have developed a new technique for bacterial identification that lowers the cost and speeds the analysis of DNA samples extracted from bacteria. The technique has significant advantages over others in use today, offering greater precision and a wider range of applications. DNA extracted from as few as 1,000 bacteria are treated with restriction enzymes that chop the DNA into fragments, the lengths of which are unique to specific bacteria. This mixture of DNA fragments is passed through a flow cytometer, and within 10 minutes the DNA fingerprint of the bacteria is known.

Implications: Using this method it is possible to distinguish between harmless strains of *E. coli* and the toxic strains that cause food poisoning. The competing technique in use today is based on pulsed field gel electrophoresis, which requires 100,000 times more cells and takes 15 hours to complete. Scientists at Los Alamos are designing an inexpensive portable version of their laboratory equipment that will be usable in hospitals and in the field.

Larson EJ, Hakovirta JR, Cai H, et al: Rapid DNA fingerprinting of pathogens by flow cytometry. Cytometry, 41(3):203-8. 2000.

Kim Y, Jett JH, Larson EJ, Penttila JR, Marrone BL, Keller RA: Bacterial fingerprinting by flow cytometry: bacterial species discrimination. Cytometry, 36(4):324-32. 1999.

Identifying the Function of Proteins

Background: A surprising result from all of the genome sequencing projects is that a third to a half of the identified genes are unique, meaning they have no homology, or sequence similarity, to previously identified genes. Although sequence homology can be misleading, it is often a useful starting point for determining the function of a gene. For genes with no sequence homology, it is necessary to develop a new technology to identify their function.

Advance: A combination of mass spectrometry, conventional light microscopy, and electron microscopy provides the needed technology to identify the function of unique genes. In a powerful demonstration project, these instruments were used to identify all of the components of the yeast nuclear pore complex, a cellular structure that allows large molecules to get in and out of the nucleus. Scientists first used traditional chromatography to separate nuclear pores from other components of the cell. Proteins isolated from the pores were then subjected to several different types of mass spectrometry to determine their sequence. This allowed the researchers to identify 174 proteins that might be associated with the nuclear pore. To distinguish the proteins that were part of the nuclear pore from those that were either found throughout the cell or were only occasionally associated with the pore, the researchers tagged the 174 proteins in an intact cell and used microscopy to observe that 134 of these molecules were located outside the membrane that contains the nuclear pore. After these "false positives" were deleted from the list of potential nuclear pore proteins, the remaining 40 proteins were tagged in an intact cell and confirmed, by electron microscopy, to be part of the nuclear pore. These experiments allowed the scientists to obtain a low-resolution structure of the pore. Prior to this work, only 6 of these 40 proteins had been assigned a function.

Implications: While the reported work was focused on determining the proteins that comprise an important cellular structure, it is straightforward to extend this technology to evaluating proteins that are present in diseased tissue. The same separation techniques could be employed to detect proteins that are more or less abundant in diseased tissue than in healthy tissue. These proteins could then be identified using mass spectrometry. Finally, the location of these proteins in the cell could be determined using light microscopy. This paradigm has the potential to identify many targets for drugs or other therapeutic measures.

Rout MP, Aitchison JD, Suprapto A, Hjertaas K, Zhao Y, Chait BT: The yeast nuclear pore complex: composition, architecture, and transport mechanism. Journal of Cell Biology, 148(4):635-51. 2000.

Most Common Type of Non-Hodgkin's Lymphoma is Actually Two Diseases, Powerful New Genetic Analysis Tool Reveals

Background: Lymphomas, which include Hodgkin's disease and non-Hodgkin's lymphoma, are the fifth most common type of cancer diagnosed and the sixth most common cause of cancer-related death in the United States. Non-Hodgkin's lymphoma is the more common of the two basic lymphoma types. This disease – often painless in its early stages – may occur in a single lymph node, a group of lymph nodes, or in another organ, and can spread to almost any part of the body, including the liver, bone marrow, and spleen. The most common type of non-Hodgkin's lymphoma is an aggressive cancer called diffuse large B-cell lymphoma (DLBCL). Forty percent of patients with a diagnosis of DLBCL are cured by standard multi-agent chemotherapy. A compelling clinical problem in the treatment of this disease has been why the remaining 60% of patients fail to respond to chemotherapy.

Advance: Using a powerful new research tool that can record the expression patterns of thousands of genes at once, NIH-supported investigators have shown that DLBCL is actually two distinct diseases. They began by using NIH's Tumor Gene Index database, a publicly available catalogue of genes expressed in various normal and cancerous cells, to identify more than 15,000 genes that are uniquely expressed in B cells. These genes were arrayed on a specialized microchip – dubbed a "Lymphochip" – along with about 3,000 other genes. Analysis of the expression patterns of the Lymphochip genes with DNA microarray technology revealed two very distinct patterns of gene expression in samples of DLCBL – a signal that the researchers were looking at two disease subtypes. A pilot study of 42 patients with DLCBL showed that individuals' responses to chemotherapy differed strikingly by disease subtype: 75% of patients with one subtype, termed "GC B-like," were alive after 5 years, whereas fewer than 25% of those with the other subtype, named "activated B-like," survived for 5 years.

Implications: This study clearly demonstrates that gene-expression profiling technology can define distinct molecular subtypes of common tumors. It points the way to a future when cancer can be diagnosed and treated based on well-characterized biological differences among tumor cells. These biological differences will enable doctors to more accurately predict how aggressive a tumor will be. An understanding of these biological differences may also allow the development of new and possibly more effective therapies tailored to the characteristics of an individual's disease.

Alizadeh AA, Eisen MB, Davis RE, et al: Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature, 403(6769):503-11. 2000.

Isolation, Purification, and Multiplication of Adult Neural Stem Cells

Background: Stem cells can reproduce themselves and can form more specialized cell types. Embryonic stem cells give rise to all cell types of the body, while stem cell types in the adult may normally have a narrower range of fates – restricted, for example, to only blood or brain cells. What these more restricted cells might be persuaded to do artificially is an open question. Perhaps an equally important question is what role do these cells play in normal brain function.

Advance: In recent years, scientists have demonstrated, surprisingly, that new nerve cells can arise in the adult human brains, so some types of stem cells are probably present. However, studies to characterize the capabilities of the proliferative cells and the signals that control them require methods to isolate, purify and multiply these cells. In the last year, scientists reached this goal using brain tissue removed during therapeutic surgery. The techniques for isolating and multiplying the human cells rely upon a variety of genetic manipulations, labeling techniques, and cell culture methods developed in studies of mouse stem cells. Taken together these studies suggest that the adult human brain harbors a complex population of stem cells that can give rise to nerve and glial cells under appropriate culture conditions.

Implications: These studies are an important step toward better understanding adult neural stem cells and developing stem cell therapies. The wide array of diseases for which stem cell therapies seem plausible arises from the remarkable versatility of these cells. Replacing lost nerve cells, replacing glial (supporting) cells, releasing growth factors, providing missing enzymes, and bridging physical obstacles to regeneration are just a sampling of the possibilities.

Beyond transplantation therapies, an equally important aspect of adult stem cell research is understanding the normal role of these cells in the brain and what the resident stem cells might be coaxed to do to repair damage. Scientists have already shown that physical exercise, stress hormones, learning, and brain trauma all affect the proliferation of brain cells, so adult neural stem cells must respond to many signaling molecules. Studying purified populations of adult human stem cells in culture, linked with experiments on stem cells within animals' brains, should foster progress in this area as well.

Roy NS, Wang S, Jiang L, et al: In vitro neurogenesis by progenitor cells isolated from the adult human hippocampus. Nature Medicine, 6(3):271-7. 2000.

Roy NS, Benraiss A, Wang S, et al: Promoter-targeted selection and isolation of neural progenitor cells from the adult human ventricular zone. Journal of Neuroscience Research, 59(3):321-31. 2000.

Wang S, Roy NS, Benraiss A, Goldman SA: Promoter-based isolation and fluorescence-activated sorting of mitotic neuronal progenitor cells from the adult mammalian ependymal/subependymal zone. Developmental Neuroscience, 22(1-2):167-76. 2000.

Roy NS, Wang S, Harrison-Restelli C, et al: Identification, isolation, and promoter-defined separation of mitotic oligodendrocyte progenitor cells from the adult human subcortical white matter. The Journal of Neuroscience, 19(22):9986-95. 1999.

Kukekov VG, Laywell ED, Suslov O, et al: Multipotent stem/progenitor cells with similar properties arise from two neurogenic regions of adult human brain. Experimental Neurology , 156(2):333-44. 1999.

Cochlear Implants

Background: The cochlear implant is the only sensory neural prosthesis in widespread clinical use. This device converts sound into electrical impulses stimulating an array of electrodes which have been surgically inserted into the inner ear, bypassing the inner ear hair cells and stimulating the auditory nerve directly.

Over 20,000 Americans with profound hearing impairment have received cochlear implants. Approximately one-half of the current cochlear implants recipients are children. The choice of a cochlear implant implies a desire on behalf of the parents to have the child fully participate in the hearing world, with spoken language skills. One of the expected benefits of cochlear implantation in children is the acquisition of spoken language. Recent data have shown that in addition to gains in spoken language, shown over the past several years, improvements in speech perception and speech production are also seen in children with cochlear implants, resulting in improved language and reading performance.

Advance: The cost effectiveness of cochlear implants in children has been questioned due to the high cost and difficulty comparing it to other medical interventions. A recent cost-utility analysis was performed for the cochlear implant use in children. Various outcome measures were used and the information was collected from parents of profoundly deaf children who had, or had not, received a cochlear implant. The cochlear implant was found to be cost-effective, providing health benefits in children at reasonable costs and results in a net savings to society.

Two new areas of investigation are being initiated: binaural implants, and a short electrode implant. Binaural implants, or implants in both ears, provide a potential avenue for better speech perception in noise, while the short electrode is being designed to be used as an adjunct in experienced, yet unsuccessful, adult hearing aid users with severe-to-profound hearing impairment.

Implications: As technology continues to provide advances in cochlear implant design, additional populations of individuals will have the potential to benefit from these remarkable devices.

Cheng AK, Rubin HR, Powe NR, et al: Cost utility analysis of the cochlear implant in children. The Journal of the American Medical Association, 284(7):850-6. 2000.

Early Identification of Hearing Impairment

Background: The NIH continues to support research focusing on early identification of hearing impairment and recently completed an initiative calling for research grants to support clinical research. In 1998, a Working Group on Early Identification of Hearing Impairment provided advice on the most important unresolved issues concerning diagnostic and intervention strategies following neonatal hearing screening. A Program Announcement was published requesting grant applications focusing on intervention strategies including: methods for fitting hearing aids, cochlear implants and other sensory aids; behavioral treatment programs; development of outcome measures to determine the benefit of intervention strategies; and studies on the efficacy of intervention. Four grants were awarded the first year, which address optimization of fitting hearing aids in infants, auditory development in early amplified children, evaluating speech therapy in toddlers, and studying the effects of very early cochlear implantation on language. This initiative has provided a funding opportunity for clinical research to resolve questions about how to intervene when hearing impairment is diagnosed in an infant. As neonatal hearing screening advances, new questions are emerging. Diligence in training an adequate pool of research clinicians remains critical to future endeavors.

Advance: One investigator has examined the relationship between age of enrollment in intervention and language outcomes in a group of deaf and hard of hearing children. Significantly better language scores were associated with early enrollment. Children enrolled before 11 months of age had stronger vocabulary and verbal reasoning skills at 5 years of age than did children enrolled at a later time. High levels of family involvement correlated with positive language outcomes, and conversely, limited family involvement was associated with significant language delays at 5 years, especially when enrollment began later in childhood.

Implications: These results provide further evidence that children will benefit when early identification of hearing loss is followed by an early intervention strategy that actively involves families.

Moeller MP: Early intervention and language development in children who are deaf and hard of hearing. Pediatrics, 106(3):E43. 2000.

Development of a High-fidelity Gene Amplification Method Allows Quantitative Measurements of Biological Functions Using Minimal Amounts of Living Tissues

Background: Biological sciences differ from more pristine disciplines such as mathematics and physics in that in the latter, theories seem to precede observation while in the former, observations struggle for a long time before they can be organized into universally accepted theories. This is because life results from a complexity of chemical and physical interactions linked by an ancestral code of about 100,000 data points called genes. Genes are the recipes of life, and complex interactions among those activated in relation to an illness are the basis of how diseases begin, progress and respond to treatment. Treatment of human diseases has traditionally targeted single genes empirically believed to be of relevance. However, this approach has been largely unsuccessful because of lack of a global view of the biological processes underlying the disease. With the completion of the human genome project, the book of life recipes is available. Yet, little is known about how genes interact with each other in sickness or in health. Microarray is a new technology that identifies genes coordinately activated during a biological process: crucial information for capturing the physiology of diseases and identify their Achilles' heel.

Advance: The large amount of RNA (the unit coding each gene) necessary for conventional microarrays has limited their utilization. To amplify RNA from minimal quantities of source material, scientists combined two powerful techniques for RNA amplification into a new method that yields 100,000-fold increase of source material while maintaining proportional the amount of each RNA type (each specific for a particular gene). Thus, the amplified RNA can be used to quantify how much of each species of RNA (gene) is present in the original material which is, in turn, proportional to the function of that gene. This technique broadens the utilization of cDNA microarrays to experimental conditions in which starting material is the limiting factor.

Implications: Direct study of living organisms is often limited by the amount of material obtainable. This is particularly true for human diseases where acquisition of biologically relevant material is limited by clinical, practical and/or ethical considerations. Similarly, research involving small organisms, embryos or fetal tissues is limited by the extraordinary small amount of RNA obtainable. This method allows the use of single cells for large studies involving microarrays. By aspirating cells from metastases (malignant growths) with a fine needle, scientists could follow the progression of events during immune treatment of patients. The scarce material obtainable by needle aspiration could be amplified to yield enough RNA to study the activation of thousands of genes in tumors at different time points using array technology. This allows the identification of 17 genes out of thousands associated with response to treatment, most of them with immune function dramatically narrowing to a defined category of biological events tumor immune responsiveness. This approach may hasten the study of biological processes associated with tissue development and with disease.

Wang E, Miller LD, Ohnmacht GA, Liu ET, Marincola FM: High-fidelity mRNA amplification for gene profiling. Nature Biotechnology, 18:457-9. 2000.

Wang E, Marincola FM: A natural history of melanoma: serial gene expression analysis'. Immunology Today, 21(12):619-23. 2000.

"Virtual Colonoscopy" One Step Closer to Routine Clinical Use

Background: Colon cancer is the second leading cause of cancer death in Americans. Colonic polyps (small growths in the wall of the colon) are thought to be precursors to colon cancer. The best test currently available for detecting and treating colonic polyps is colonoscopy. However, it is expensive and uncomfortable and only a fraction of Americans have it done despite health guidelines recommending it after age 50. Radiologists have spent the past six years investigating a new way to diagnose colonic polyps, called computed tomography (CT) colonography or "virtual colonoscopy." In virtual colonoscopy, the patient undergoes a CT scan of the abdomen and pelvis to look at the colon. A small tube is placed into the patient's rectum and the colon is distended with air. The test takes only a few minutes. Surveys have shown that many patients prefer virtual colonoscopy over conventional colonoscopy.

Advance: Researchers at NIH have taken the first step towards developing a computer algorithm that analyzes the CT scan and locates possible polyps automatically. The computer algorithm creates a three-dimensional model of the patient's colon. Then the computer inspects each of hundreds of thousands of points on the model of the colon to look for small protrusions which could represent a polyp. Using various criteria, the algorithm prunes the list of possible polyps to a manageable few and reports these to the physician. Researchers have proven feasibility of this algorithm on a computer simulation of colonic polyps derived from human virtual colonoscopy data. They are currently validating the algorithm on virtual colonoscopy studies of patients with known polyps.

Implications: Virtual colonoscopy needs to be cost-effective and accurate if it is to be widely used. One major barrier to cost-effectiveness is that it takes a physician anywhere from 10 to 60 minutes to interpret a virtual colonoscopy study. One major barrier to accuracy is perceptive error (the physician does not locate a polyp that can be seen in retrospect). A computer algorithm such as the one we have described could address these barriers to cost-effectiveness and accuracy. The algorithm would prescreen the colon for polyps thereby reducing interpretation time and increasing accuracy of polyp detection. The goal: a test that patients prefer and that reduces suffering from colon cancer.

Summers RM, Beaulieu CF, Pusanik LM, et al. An automated polyp detector for CT colonography: feasibility study. Radiology, 216(1):284-90. 2000.

Lumke HU, Vanier MW, Inamura K, et al: Computer assisted radiology and surgery(CARS). San Francisco, CA: Elsevier Science, 785-789. 2000.

Magnetic Resonance Angiography Displays Small Vessels Without Invasive Catheterizations

Background: Arteriograms are radiologic studies in which direct punctures of the arteries are performed to insert catheters through which contrast media is injected at high rates. Such studies are invasive and can cause vascular injury. A new method of displaying arteries based on magnetic resonance imaging, known as MRA for magnetic resonance angiography, has been developed which requires only a minimally invasive intravenous injection. This study, which can be performed within minutes, is beginning to replace conventional arteriography of large vessels like the aorta; however, it has not yet proven useful for identifying smaller vessels.

Advance: Researchers at NIH have developed a computer program, which automatically traces the course of a vessel on a MRA from its origin to its termination. Using this technique, dubbed "skeletonization" the user defines the origin and end of the vessel and the algorithm identifies the center of the vessel as a colorized three-dimensional model, allowing the physician to see the course of smaller vessels within structures such as the liver, spleen and legs. Individual vessels can be identified by their differing colors.

Implications: Identification and display of smaller arteries within the body will allow magnetic resonance angiography to be used in place of conventional angiography in almost all cases. Compared to existing methods of image display, skeletonization offers a quick, comprehensive synopsis of vascular anatomy. This advance opens new opportunities for additional methods of vessel display incorporating not only anatomic information but also flow information and identification of atherosclerotic plaques or hardening of the arteries, which leads to blockage of vessels. A family of such computer programs, which enhance vessel display, entitled "Vesselize," is being developed for desktop personal computers which will allow this technology to disseminate into general use.

Choyke PL, Yim P, Marcos H, Ho VB, Mullick R, Summers RM: Hepatic MR Angiography: A Multiobserver Comparison of Visualization Methods. American Journal of Roentgenology, 176.February 2001.

Yim PJ, Mullick R, Choyke PL. Vesselize: a system for analysis and visualization of magnetic resonance angiography. Journal of Digital Imaging, 13(2):219-20. 2000.

Zeh H, Choyke PL, Alexander HR, et al: Gadolinium enhanced 3D MRA prior to isolated hepatic perfusion for metastases. Journal of Computer Assisted Tomo graphy, 23(5):664-9. 1999.

Improved Integration of Data from Multiple Gene Maps Leads to Better Localization of Human Genetic Markers

Background: Radiation-Hybrid (RH) maps have been vital to the human genome sequencing project by providing gene sequencers with landmarks to guide their efforts. Several independent RH maps have been produced, such as the GB4 and G3 maps used in GeneMap99, which included more than 35,000 human gene-based markers constructed by the International Radiation Hybrid Mapping Consortium and made available for searching by the NIH. These various maps, however, use different markers and do not always place common markers in the same order on the human sequence. A composite map of data from several RH maps was needed to provide sequence landmarks at a higher density and of greater accuracy than any single RH map.

Advance: Radiation-Hybrid maps are linear sequences of DNA markers that are oriented with respect to one another using their RH vectors. The smaller the difference between two RH vectors, the nearer the corresponding markers are to one another in the DNA sequence. The linear sequence of markers on a map is estimated by choosing an order which minimizes the sum of the differences between the RH vectors of each pair of consecutive markers on the map. Classical ordering algorithms produce a "good" ordering but this is not guaranteed to be the best and improvements are possible. The current advance was achieved by combining data from more than one independent RH marker set and applying an algorithm called CONCORD, an adaptation of an algorithm commonly used to solve a mathematical conundrum known as the "Traveling Salesman" problem. The CONCORD algorithm produces a combined RH map having a higher density of markers and a lower average RH-vector distance than that of any single RH map.

Implications: The higher quality, higher density RH-map produced using the CONCORD method will enable the remaining portions of the human genome to be placed on the existing human gene map with greater accuracy. In addition, the combined map can be used to precisely identify genes known to be associated with RH markers. This will be of great help to disease-gene hunters.

Agarwala R, Applegate DL, Maglott D, Schuler GD, Schaffer AA: A fast and scalable radiation-hybrid map construction and integration strategy. Genome Research, 10(3):350-64. 2000.

LocusLink and RefSeq Databases Simplify Integrated Retrieval of Genomic Information

Background: As the number of human gene sequences in the NIH's GenBank DNA sequence database continues to grow, it becomes increasingly difficult to choose a representative sequence for a particular gene. Because GenBank is a comprehensive and archival data repository, there can be many duplicate or overlapping sequences in the database for a given gene, each submitted by a different laboratory. The number of different database resources containing discrete pieces of genomic information is also growing rapidly, making it difficult for scientists to form a complete picture of what is known about a gene without consulting a diverse set of resources on an ad hoc basis.

Advance: NIH has developed two integrated resources to address these problems – RefSeq to provide a single standard reference sequence for each human gene, and LocusLink to serve as a central hub that links the diverse information resources pertaining to each gene. Together they provide greatly improved methods for organizing and disseminating genomic information.

RefSeq is a database containing curated human and non-human mRNA sequences, each intended to be the best representative sequence for a particular gene. To produce a RefSeq entry, the most complete version of a gene sequence is chosen from GenBank to serve as an initial provisional sequence to represent that gene. The provisional sequence may then be edited and extended with data from other sources to produce a standard reference sequence of the highest quality and greatest length possible. A companion protein reference sequence is also created to represent the protein product of the gene. The RefSeq record is then further enhanced with additional biological annotation not found in the original GenBank record. The result is a compact database of highly reliable sequences with superior annotation. RefSeq is available at: http://www.ncbi.nlm.nih.gov/LocusLink/refseq.html

LocusLink is a database that draws together information on human genes, as well as those from other organism. A LocusLink gene report includes information on reference sequences, map locations, synonyms, associated disease names and phenotypes, clinical synopses, three-dimensional structures for protein products, and more. The LocusLink database can be queried using almost any item if information contained within it. LocusLink is available at: http://www.ncbi.nlm.nih.gov/LocusLink/.

Implications: Arriving at a representative sequence and a comprehensive collection of information relating to a single human gene has now been greatly simplified. With the advent of RefSeq, researchers no longer have to examine and choose between a dozen nearly identical gene sequences before beginning a sequence analysis project. By using LocusLink, researchers can easily access the diverse array of information pertaining to a disease-associated gene, such as map locations, marker linkages, clinical synopses, and protein modeling templates for drug design. The integration of resources in LocusLink will not only save the research community time but will present researchers with a broad band of information, much of which they might otherwise be unaware. The combination of RefSeq and LocusLink will accelerate the rate at which the genes comprising the human genome are understood.

Pruitt KD, Katz KS, Sicotte H, Maglott DR: Introducing RefSeq and locuslink: curated human genome resources at the NCBI. Trends in Genetics, 16(1):44-7. 2000.

Maglott DR, Katz KS, Sicotte H, Pruitt KD: NCBI's locuslink and RefSeq. Nucleic Acids Research, 28(1):126-8. 2000.

ClinicalTrials.gov

Background: The FDA Modernization Act of 1997 required the Department of Health and Human Services, through the NIH, to establish a registry of clinical trials for both federally and privately funded trials "of experimental treatments for serious or life-threatening diseases and conditions", thereby broadening the public's access to information about clinical trials to a broad range of diseases. (Section 113, Information Program on Clinical Trials for Serious or Life-threatening diseases, Food and Drug Administration Modernization Act of 1997, Public Law 105-115, 105th Congress.).

Advance: *ClinicalTrials.gov* was launched on February 29, 2000. The system currently contains information on more than 5,000, primarily NIH sponsored, clinical trials. *ClinicalTrials.gov* provides patients, families and members of the public easy access to information about the location of clinical trials, their design and purpose, criteria for participation, and, in many cases, further information about the disease and treatment under study. There are also contacts to individuals responsible for recruiting participants for each study. Other information in the database that may help a patient decide whether to enroll in a particular trial includes the research study design, the phase of the trial, the disease or condition, and the particular drug or therapy under study.

The system is freely available over the World Wide Web, and patients and others are able to search and browse the system in a variety of ways. They may choose to browse the system by looking, for example, at a list of disease names, or at a list of sponsors. Alternatively, they may prefer to interact with the system by entering a word or a phrase that the system searches across the entire database. Or, they may wish to narrow down their searches by looking for trials on their specific disease that are being conducted in a particular location.

ClinicalTrials.gov is linked to the online medical literature and other related health resources. If a study has been completed and the investigators have published the results of their research, then these link to NIH's MEDLINE database, which contains almost 10 million references to articles in some 4,000 journals in the health sciences. *ClinicalTrials.gov* is also closely linked to NIH's consumer health site, MEDLINE*plus*, which makes available information on more than 400 health topics. NIH has many other online sources of health information, including dictionaries, reviews of recent research results, and other educational materials. Links to these resources may be helpful to patients and others who would like to get a better understanding of what is involved in a clinical trial, what the standard therapy for a particular disease is, or even to better understand the disease itself.

Implications: Clinical trials are medical research studies that seek to evaluate the safety and effectiveness of new drugs, medical procedures, or other means of treating, diagnosing, or preventing diseases. This type of research helps investigators learn how different people respond to medications or other therapeutic approaches, and such investigations may lead to new or improved treatments. Trials are conducted when there is no proven treatment for a specific disease, or to test which treatment

works best for a particular disease or condition. Patients may want to participate in a clinical trial for a variety of reasons. Most often they are hopeful that the experimental treatment will help them. Many patients who participate in clinical trials believe in the importance of medical research, and, after weighing the benefits and possible risks of participating in a trial, decide to participate because this allows them to contribute personally to a possible cure or better treatment for their disease.

ClinicalTrials.gov provides a single place for patients and others to come to find information about clinical trials. Such a system has never before been available. People now have a central, easily accessible source of information for clinical trials no matter where the trials are being conducted, what the particular drugs or diseases under study are, or who is sponsoring the trials.

McCray AT, Ide NC: Design and implementation of a national clinical trials registry. Journal of the American Medical Informatics Association, 7(3):313-23. 2000.

http://www.nih.gov/health/clinicaltrials

Profiles in Science

Background: The NIH is making the archival collections of pioneering biomedical scientists of the 20th century available on its *Profiles in Science* Web site. The site promotes the use of the Internet for research and teaching in the history of biomedical science. The collections have been donated to NIH and contain published and unpublished items, including books, journal volumes, pamphlets, diaries, letters, manuscripts, photographs, audiotapes, video clips, and other materials.

Advance: Each digital collection on the *Profiles in Science* site consists of two major parts. First is an exhibit, composed of introductory narratives on the scientist's life and work and between 60 and 75 noteworthy documents (text, audiotapes, video clips and photographs). The exhibit is particularly designed for students and those with little background in science. The second part consists of additional documents from the scientists' papers, available through a search engine and in alphabetical and chronological "views." In addition, for the Joshua Lederberg collection, the donor has provided commentary in the form of annotations to individual documents.

The site was launched in September 1998 with a selection of papers devoted to the life and work of Oswald T. Avery (1877- 1955). Avery was a founder of modern molecular biology and the first person to establish that genetic information is encoded in DNA.

In March of 1999, the site was expanded with the release of a portion of the extensive Joshua Lederberg (b. 1925) collection. This digital archive reflects Lederberg's work in bacterial genetics, artificial intelligence, exobiology, biological warfare, and public health, as well as his role as a public figure in issues of science and society. Lederberg received the Nobel Prize in 1958 for "his discoveries concerning genetic recombination and the organization of the genetic material of bacteria."

In November 1999, a selection from the papers of Martin Rodbell (1925-1998) was posted. Rodbell was a biochemist and molecular endocrinologist who shared the Nobel Prize in 1994 for the discovery of "G-proteins and the role of these proteins in signal transduction in cells."

In May 2000, the Julius Axelrod (b. 1912) collection was made available. Axelrod is a pharmacologist and neuroscientist who shared the 1970 Nobel Prize for the discoveries "concerning the humoral transmittors in the nerve terminals and the mechanism for their storage, release and inactivation." New collections are continually being processed and will be added as they are ready.

Implications: *The Profiles in Science* site is a growing resource for students, educators, and researchers. The site makes the successes of science readily accessible to anyone who is interested in the scientific process and in the people who have dedicated their lives to scientific discovery. The collections are particularly strong in the areas of cellular biology, genetics, and biochemistry, but also reflect issues in such areas as health and medical research policy, the application of computers in medicine, science education, and the search for extraterrestrial life.

http://profiles.nlm.nih.gov/

Development of a New Technique for Examining Different Structures in the Eye

Background: Throughout the history of studies in biological science and medicine, imaging of organs and tissues has been essential for diagnostic and therapeutic purposes. The development of new methods for visualizing normal and pathological tissues is critical for advancing our understanding and treatment of diseases. Optical Coherence Tomography (OCT) is an imaging technique that is analogous to ultrasound because two-dimensional images of structures in the body are obtained non-invasively. Unlike ultrasound, which uses sound waves, OCT uses infrared light to probe the structures of interest. The use of light rather than sound permits acquisition of clearer images with excellent spatial resolution (~ten micrometers). The high resolving power of OCT and its versatility permit its adaptation to a wide variety of problems. It is particularly well suited to applications in ophthalmology because it is non-contact, adaptable to ophthalmic instrumentation, and permits imaging structures at different depths within the eye.
Advance: OCT is a relatively new imaging technique, and a great deal of developmental work continues. Scientists have produced a new signal processing approach called Color Doppler OCT that permits acquisition of images at higher speeds combined with greater resolution than previously achieved. The increased acquisition rate will eventually permit visualizing moving structures without any delay in processing. In addition, a novel hand-held OCT probe was developed that is small and easily manipulated by the examiner. When combined with the higher speed of image acquisition, this instrument could image the front of the eye in real-time. Scientists are now able to see various detailed features of the cornea, the clear tissue in the front of the eye. These include the outermost corneal cell layer (the epithelium) and deeper layers of the cornea below the epithelium.

Implications: Real-time, high resolution OCT imaging of the structures in the front of the eye may be useful in evaluating a variety of structural causes of glaucoma overcoming the technical barriers that have heretofore existed. Similarly, high speed data acquisition combined with the outstanding spatial resolution of OCT will enable imaging of the retinal microcirculation nearly in real-time. This will be of tremendous value in diagnosing a variety of macular diseases in a non-invasive manner. For example, in diabetic retinopathy, pathological blood vessels impinge on the retina, eventually destroying vision. OCT will permit earlier detection and treatment yielding a better outcome and improved sight over the long-term.

Yazdanfar S, Rollins AM, Izatt JA: Imaging and velocimetry of human retinal circulation using color doppler optical coherence tomography. Optics Letters, 25:1448-50. 2000.

Development of a Multimedia Educational Tool About the Human Genome and Genetics for Students and the Public

Background: It is well documented that scientific literacy, and genetic literacy in particular, are low in the U.S. In an effort to make the Human Genome Project and genetic research more accessible and interesting, NIH is creating a multimedia kit, entitled "*The Human Genome: Exploring our Molecular Selves*." Its development was driven by the motivation to build genetic literacy and to share the excitement and awe of the human genome.

Advance: The kit is designed to educate, engage, and excite users about the human genome and genetics. The primary target audience is high school students though a much broader use with college students, voluntary health organizations and the general public is anticipated. This project is being co-sponsored by the NIH, Department of Energy, Howard Hughes Medical Institute, Pharmaceutical Research and Manufacturers of America (PhRMA), and the journals Science and Nature. The kit will be released with publication of the human sequence data.

The kit includes a multimedia CD-ROM, 15-minute video documentary, wall poster of the genome, and an informational brochure. The video documentary shares the excitement of unlocking the secrets of the genome by tracing the development, evolution and impact of the Human Genome Project and genomics research through the words of leaders in human genome research. The CD-ROM consists of six elements:

- A 3-D molecular animation, rich images and "state-of-the-art" visual techniques;
- An Interactive, comprehensive timeline with over 90 key events and milestones in genetics;
- An animated segment with narration on "how to sequence a genome;"
- A variety of classroom activities;
- A segment addressing ethical and societal issues; and
- A new edition of NHGRI's talking glossary of genetic terms.

Implications: Knowledge of the human genome is increasing exponentially. The application of this knowledge to improve health and medicine and expand our understanding of ourselves as a species and as communities, families, and individuals will be profound. To maximize the benefits and to ensure informed public and personal decision making about science, medicine and policy, the public must understand basic concepts of genomics and genetics.

Once released, the kit will be accessible at: http://www.nhgri.nih.gov/educationkit

Accumulation of Cellular Iron: A Likely Culprit in Degenerative Diseases of the Nervous System

Background: Iron is necessary to transport oxygen throughout the body and is the primary determinant of how much oxygen reaches and is used by all the cells. However, iron levels have to be maintained in a delicate balance. While too little iron leads to reduced availability of oxygen throughout the body, too much iron is toxic to the cells and may cause or aggravate a wide variety of diseases. For example, iron toxicity is thought to be a factor in the development of diseases such as atherosclerosis or chronic hepatitis. The NIH has been a leader in discovering how iron is taken into the cells and how it is sustained at constant levels.

Iron uptake and storage is controlled by a number of proteins. Scientists have identified many of the proteins involved in maintaining normal iron levels and have increased their understanding of how these proteins' functions are regulated. Clinicians also know that iron accumulates in nerve cells that are damaged, and researchers have hypothesized that excess iron might play a role in such diseases as Parkinson's and Alzheimer's Diseases. However, it is difficult to distinguish cause and effect, i.e., is the excess iron the cause or the result of the disease process?

Advance: Recently, scientists tested the effects of inactivating of one of the proteins that is responsible for iron metabolism in a genetically-altered mouse model. The result was mice that displayed progressive symptoms of neurological degeneration that closely paralleled those of a human condition known as Multiple System Atrophy (MSA). These mice also showed abnormal accumulation of iron in neurons and other nerve cells within specific locations in the central nervous system, similar to the increased iron accumulation reported in the central nervous system in humans with MSA.

Implications: The fact that iron accumulates in nerve cells long before the cells die, strongly suggests that iron is the cause rather than the result of impaired neural function. This finding could have implications for diagnosing and treating a wide range of diseases of aging. Furthermore if the inability to regulate cellular iron levels is indeed involved in degenerative diseases of the nervous system, then this mouse model will be very useful in developing therapies to reduce iron levels. Drugs that bind excess iron and render it harmless are already in use to treat iron overload in the liver and heart. Efforts are under way to design similar drugs that can cross the blood-brain barrier to prevent neuronal damage. This animal model will be very valuable for testing such drugs, which may ultimately provide a treatment for many devastating progressive neurological diseases.

LaVaute T, Smith S, Land W, et al: Targeted deletion of iron regulatory protein 2 causes iron overload and neurodegenerative disease in mice. Nature Genetics (in press).

A Simple, Inexpensive Method for Determining Prognosis in B Cell Lymphocytic Leukemia

Background: The leukemias are cancers that affect various types of blood cells. Some types of leukemia attack blood cells called lymphocytes, which play an important role in the body's immune response. Leukemias are classified according to their clinical course as either acute or chronic. One of the chronic lymphocytic leukemias, B cell chronic lymphocytic leukemia (B-CLL), is the most common type of leukemia in the Western world. About 7,500 individuals develop and 5,000 die from B-CLL each year. The disease especially strikes older adults, increasing in incidence with each decade of life after 40, and affects nearly twice as many men as women Once a patient has developed B-CLL, the disease may take any of a number of different courses. Some patients survive for long periods without needing definitive therapy, whereas others decline rapidly and die despite receiving aggressive treatment. Sophisticated genetic techniques can be used to determine the extent of the disease, information that is useful in determining the clinical course it is likely to take. These genetic tests, however, are labor intensive, expensive, and not widely available. Clinical diagnosis, treatment, and prognosis for B-CLL would be greatly facilitated by a simpler technique that could be used relatively easily and inexpensively in most hospitals. Now NIH-supported researchers have developed a technique that may fit the bill.

Advance: Currently, B-CLL can be staged by using genetic techniques to determine the presence or absence of mutations in a part of the gene known as the immunoglobulin variable (V) region. Patients whose tumor cells do not have mutations in this region have a poor prognosis, whereas the outlook is better for those with mutated V region genes. Making this determination, however, requires highly skilled technicians and sophisticated equipment. The investigators used flow cytometry, a relatively simple technique available in many hospitals, to look at proteins on the surfaces of cancerous B cells. What they discovered concerned the CD38 protein, which resides on the surfaces of B-CLL tumor cells. The investigators found that the density of this protein on cell surfaces was directly proportional to the degree of V region gene mutations – and the clinical course of the patient's disease.

Implications: Instead of looking for genetic mutations to determine clinical stage of B-CLL, flow cytometry can be used to determine the density of CD38 protein on cell surfaces. In turn, CD38 density serves as an indicator of the degree of mutation. This assay, which can be performed simply and relatively inexpensively in most hospitals equipped with a flow cytometer, holds promise for enabling clinicians to more accurately assess the stage of B-CLL and its prognosis, thereby aiding in tailoring more effective treatments for these patients.

Damle RN, Wasil T, Fais F, et al: Ig V gene mutation status and CD38 expression as novel prognostic indicators in chronic lymphocytic leukemia. Blood, 94(6):1840-7. 1999.

How Cells Engineer Connective Tissue Matrix

Background: Cells in our body's tissues are in contact with neighboring cells and the connective tissue matrix that is layered between cells. Made up of proteins, this mesh-like matrix provides a structure in which cells can move, grow, and organize into tissues. But little is known about how cells produce and organize protein fibers to create this connective tissue matrix. One of the matrix proteins, called fibronectin, plays a crucial role in normal embryonic development and wound healing. However, it also contributes to the development of fibrosis, or excess connective tissue, and scar tissue. Additionally, fibronectin is thought to be involved in the spread of cancer. Because of its many roles in the body, it is important to understand how this protein functions.

Advance: NIH researchers have discovered the biological machinery that cells use to create fibronectin fibers. The machinery uses an escalator-like movement that works like this: fibronectin receptors called integrins move along filament "tracks" (inside the cell) using another protein called tensin. The motion of the integrins and tensins appears to stretch fibronectin (on the outer surface of the cell) into long fibers. These fibers then become organized into a three-dimensional matrix. The scientists also found that several different drugs halt or accelerate the integrin motion and the subsequent creation of fibronectin fibers.

Implications: Knowing how fibronection fibers are created, scientists may one day be able to control or mimic the process. This could aid in tissue engineering procedures, provide new approaches for reducing fibrosis and scarring, and lead to the development of new techniques for stopping the spread of cancer at the molecular level.

Pankov R, Cuikierman E, Katz BZ, et al: Integrin dynamics and matrix assembly: tensin-dependent translocation of $a_5ß_1$ integrins promotes early fibronectin fibrillogenesis. The Journal of Cell Biology, 148(5):1075-90. 2000.

Zamir E, Katz M, Posan Y, et al: Dynamics and segregation of cell-matrix adhesions in cultured fibroblasts. Nature Cell Biology, 2(4):191-6. 2000.

New Genetic System for Following Cranial Neural Crest Cells in Mice

Background: Neural crest cells are multipotential stem cells (individual stem cells that can clonally produce all of the different cell types and are self-renewing) derived from the neural tube, that is, the epithelial tube in the embryo that forms the central nervous system. They migrate extensively during vertebrate development and give rise to a wide variety of cell and tissue types. In mammals, studies of the final destiny of migrating neural crest cells have been hindered by the lack of an effective marker that would permit clear identification of these cells as they migrate and develop.

Advance: A neural crest reporter gene (Wnt1) has allowed scientists to track neural crest cells in mice during part of their migration. Researchers have now developed a two-component genetic system which includes modifying Wnt1 and adding a second gene (ROSA26) which is turned on after Wnt1 switches off. This has allowed them to follow the migration and differentiation of mouse cranial neural crest cells to their final destination. They found that cranial neural crest cells contribute to the formation of teeth, other craniofacial structures, and elements of the cardiovascular system.

Implications: This new model provides a valuable tool that will lead to a more comprehensive understanding of the contribution of the mammalian neural crest to craniofacial development and its role in abnormal embryogenesis. It should lead to increased knowledge of a range of developmental disorders and permit better evaluations of the effect of potential corrective measures, including gene therapy.

Chai Y, Jiang X, Ito Y, et al: Fate of the mammalian cranial neural crest during tooth and mandibular morphogenesis. Development, 127(8):1671-9. 2000.

Early Warning System is Discovered to Fight Infections and Cancer

Background: The innate components of the immune system employ defense mechanisms that act quickly, without differentiating among the infectious agents and tumor cells they attack. In contrast, the adaptive components of the immune system respond to specific infectious agents or tumors. The adaptive components of the immune system respond less rapidly than the innate components because adaptive responses require proliferation of the cells (types of T and B cells) that target specific infectious agents. Such specificity is controlled by the interaction between major histocompatibility complex (MHC) molecules and T cells. These interactions activate the immune system to fight the existing infection, as well as to respond more quickly to a subsequent infection with that particular agent. In comparison, the methods by which the innate components of the immune system respond to infected cells are less understood.

Advance: NIH-supported researchers have now identified two key molecules, MICA and MICB, which alert the innate immune system to respond against infectious agents and tumor cells. Using a method, called X-ray crystallography, for determining the three-dimensional structure of proteins, the researchers determined that MICA and MICB are very similar in structure to MHC molecules. However, the MICA and MICB molecules play a different role in controlling infections and cancer than MHC does. Unlike MHC molecules, which activate T cells, MICA and MICB serve as "tags" that identify infected or cancerous cells for destruction by natural killer (NK) cells. More precisely, the NIH-investigators demonstrated that MICA and MICB bind to a molecule on the NK cell surface called NKG2D, which triggers the NK cell to attack its target.

Additional findings from the same team of investigators demonstrated that MICA and MICB are displayed on the surface of cells that are infected with cytomegalovirus (CMV), a major cause of illness in transplant recipients. Given that CMV is known to hinder T cell responses by inhibiting production of MHC molecules, the presence of MICA and MICB may allow the innate response to compensate for the suppression of the adaptive T cell response.

Implications: These studies provide significant information about MICA and MICB molecules and the role they play in the innate immune response to infections and tumors. In addition, because increasing evidence suggests that the innate components of the immune system can enhance the activity of the adaptive components of the immune system, vaccines may be designed and developed based on MICA and MICB targets.

Bauer S, Groh V, Wu J, et al: Activation of NK cells and T cells by NKG2D, a receptor for stress-inducible MICA. Science, 285(5428):727-9. 1999.

Wu J, Song Y, Bakker ABH, Bauer S, Spies T, Lanier LL, Phillips JH: An activating immunoreceptor complex formed by NKG2D and DAP10. Science, 285(5428):730-2. 1999.

Li P, Willie ST, Bauer S, Morris DL, Spies T, Strong RK: Crystal structure of the MHC class 1 homolog MIC-A, a ?d T cell ligand. Immunity, 10(5):577-84. 1999.

Groh V, Rhinehart R, Secrist H, Bauer S, Grabstein KH, Spies T: Broad tumor-associated expression and recognition by tumor-derived ?d T cells of MICA and MICB. The Proceedings of the National Academy of Sciences, 96(12):6879-84. 1999.

Imaging Technique Reveals Changes in Brain Structure

Background: Both human development and degenerative disease processes are known to affect the volume of brain substructures. But detecting, tracking, and quantifying these structural changes have proven difficult, in part because observations must be made throughout the entire volume of the brain, and they must be made at different times. Until recently, scientists have been unable to obtain detailed three-dimensional maps of these dynamic growth processes.

Advance: Researchers at UCLA have created detailed three-dimensional images that map growth patterns in the developing human brain over time. A new tensor mapping strategy allows greater spatial detail and sensitivity than was previously obtainable. The researchers discovered that different parts of the brain grow at markedly different rates during the development of a normal child. Between ages of 3 and 6, peak growth rates were observed in an area of the brain responsible for mental vigilance and the planning of new actions. Older children displayed fastest growth in a region of the brain related to spatial association and language function. Growth in these areas slowed considerably once children were 11 to 15 years old. This temporal pattern coincides with the ending of a well-known critical period for learning language. The ability to learn new languages declines rapidly after the age of 12, as does the ability to recover language function if linguistic areas in one brain hemisphere are surgically removed.

Implications: This group of researchers recently found that the same areas of the brain that grow fastest in children also degenerate fastest during the early stages of Alzheimer's disease. The sensitivity of the new experimental protocol may offer advantages in tracking the effects of various treatments for Alzheimer's disease. This approach can also be extended to evaluate the effects of therapies for other brain disorders, such as dementia and brain tumors.

Thompson PM, Gledd JN, Woods RP, MacDonald D, Evans AC, Toga AW: Growth patterns in the developing brain detected by using continuum mechanical tensor maps. Nature, 404(6774):190-3. 2000.

Advances in Ocular Kinematics

Background: Ocular saccades are coordinated movements of the eyes. Saccades occur, for example, when we are reading, watching TV, or scanning the environment. This coordinated movement is accomplished primarily by two pairs of extra ocular muscles (EOM), the medial and lateral rectus, that are attached to the globe. Neural signals from the brainstem activate these muscles causing activity in these agonist/antagonist muscle pairs which results in conjugate eye movement and maintenance of binocular vision. In visual disorders such as strabismus ("crossed eyes"), the two eyes do not function as a unit and the result is a loss of binocular vision. Surgical and chemical approaches to correcting strabismus act by affecting the extraocular muscles. These treatments are often effective but may have unintended effects. Therefore, a better understanding of the function of the extraocular muscles could lead to improvements in therapy.

Advance: An NIH grantee has developed the "active pulley hypothesis" to explain currently unresolved issues in the kinematics of eye movement. Magnetic Resonance Imaging (MRI) data from normal subjects has permitted a high-resolution picture of the globe and associated extraocular muscle and connective tissue. The data clearly show that EOMs have different insertion points. The globular fibers of EOM terminate on the sclera of the globe while the orbital layer terminates not on the globe itself, but on connective tissue in the orbit. This work demonstrates that the globular layer rotates the eye while the orbital layer functions as a pulley to influence the axis of rotation of the EOM. Importantly, the orbital and globular muscle fiber types are anatomically, physiologically and biochemically distinct. This suggests significant functional consequences for orbital and globular insertion on eye movement.

Implications: Treatment of strabismus probably affects the action of orbital EOM layers on their connective tissue pulleys. The identification of this organization of EOM fibers has important implications for mathematical models of rotation properties in the eye and will likely impact surgical approaches. Any pulley manipulation which compromises the orderly relationship between eye orientation and the rotational axes of the EOMS would compromise neural control of eye movement. This would be expected to produce at least dynamic binocular misalignments in gaze.

Demer JL, Oh SY, Poukens V: Evidence for active control of rectus extraocular muscle pulleys. Investigative Ophthalmology and Visual Science, 41:1280-90. 2000.

SCIENCE CAPSULES

Flourescent Brain Cells Help Track Master Reproductive Hormone. Gonadotropin releasing hormone (GnRH), produced by cells in the brain, is the master hormone that starts a sequence of cellular events, which control male and female reproduction. Until recently, however, scientists have been unable to throughly study GnRH, because it is produced by a small number of cells that are widely scattered throughout the hypothalamus region of the brain. Now, an investigator has developed a strain of mice whose GnRH neurons glow with a fluorescent jellyfish protein, providing researchers with a visible target to track the production and release of this key reproductive hormone in living animals.

Suter KJ, Song WJ, Sampson TL, Wuarin JP, Saunders JT, Dudek FE, Moenter SM: Genetic targeting of green fluorescent protein to gonadotropin-releasing hormone neurons: characterization of whole-cell electrophysiological properties and morphology. Endocrinology, 141(1):412-9. 2000.

Detailed Images Achieved with New Magnetic Resonance Imaging (MRI) Technique. Scientists recently developed a new MRI technique that produces the clearest, most detailed images yet obtained of coronary arteries in living individuals. The non-invasive technique "blacks out" the motion of blood flow that obscures the image of the artery wall in ordinary MRI pictures. By providing detail of the composition of atherosclerotic plaques (the cholesterol-rich deposits narrowing diseased arteries), the technique offers the potential to identify "vulnerable" plaques – those plaques most likely to rupture and lead to sudden clot formation and an acute heart attack. This would enable physicians to diagnose individuals at risk for a heart attack before they suffer one..

Fayad ZA, Fuster V, Fallon JT, et al: Noninvasive in vivo human coronary artery lumen and wall imaging using black-blood magnetic resonance imaging. Circulation, 102(5):506-10. 2000.

Improved Prospects for Gene Therapy Requiring Large Genes. Some genetic diseases, such as hemophilia A, have posed problems for gene therapy because the corrective gene is too large to fit inside the modified viruses that can be used to carry it into a cell. But researchers have now found a way to split large genes into two parts and package them into complementary pairs of a virus known as adeno-associated virus (AAV). The corrective gene is divided into one part that carries the genetic code for the protein needed to treat the disease and a second part that contains all the genetic sequences necessary to regulate that protein's production inside the target cell. The two parts "join hands" inside the cell and permit high level production of the therapeutic protein. Thanks to this breakthrough, AAV will be able to deliver therapeutic genes for diseases previously believed to be untreatable by gene therapy.

Duan D, Yue Y, Yan Z, Engelhardt JF: A new dual-vector approach to enhance recombinant adeno-associated virus-mediated expression through intermolecular *cis* activation. Nature Medicine, 6(5):595-8. 2000.

Nakai H, Storm TA, Kay MA: Increasing the size of rAAV-mediated expression cassettes in vivo by intermolecular joining of two complementary vectors. Nature Biotechnology, 18(5): 27-32. 2000.

Yan Z, Zhang Y, Duan D, Engelhardt JF: Trans-splicing vectors expand the utility of adeno-associated virus for gene therapy. Proceedings of the National Academy of Science , 97(12):6716-21. 2000.

New Method to More Quickly Study Cellular Regulatory Processes. The most common and probably the most important way that protein activity is regulated is through a process known as phosphorylation. Phosphorylation involves adding phosphorus molecules to proteins, thus altering their shape and ability to perform. To better understand these protein regulatory processes, it is necessary to identify the amino acids on the protein that are phosphorylated. A new technique has been developed that combines metal ion affinity and direct analysis by mass spectrometry. It is useful for detecting phosphorylated peptides and can be used for analyzing extremely complicated mixtures. It eliminates the need for use of radioisotopes and for prior protein separations using high-performance liquid chromatography (HPLC), thus enabling scientists to more quickly analysis critical steps in regulation.

Zhou W, Merrick BA, Khaledi MG, Tomer KB: Detection and sequencing of phosphopeptides affinity-bound to immobilized metal ion beads by matrix-assisted laser desorption/ionization mass spectrometry. Journal of American Society for Mass Spectrometry, 11(4):273-82. 2000.

Automatable Method for Protein Identification. Identification of proteins is becoming increasingly important as attention turns from characterizing the genome to understanding the functional consequences of gene activity, e.g., what proteins are synthesized and how do they perform their biological function. Scientists have developed a procedure for the identification of proteins separated by gel electrophoresis without extensive reduction and chemical derivatization using a low flow rate continuous introduction system and tandem mass spectrometry. Omitting the reduction/derivatization step reduces sample contamination and sample loss.

Borchers C, Peter JF, Hall MC, Kunkel TA, Tomer KB: Identification of in gel-digested proteins by complementary peptide-mass fingerprinting and tandem mass spectrometry data obtained on an electrospray ionization quadrupole time of flight mass spectrometer. Analytical Chemistry, 72(6):1163-8. 2000.

Efficient Production of Dopamine Nerve Cells. The major symptoms of Parkinson's disease reflect the progressive death of a specific set of nerve cells in a brain area called substantia nigra. Nerve cells in this region send fibers to another part of the brain (the striatum) where they release the neurotransmitter dopamine. Surgically replacing these cells has shown therapeutic promise in animal models of Parkinson's disease and (using tissue implants) in human patients. However, lack of access to large numbers of dopamine neurons limits the development of cell replacement therapies. NIH

intramural scientists have now developed a method to coax mouse embryonic stem cells to produce large numbers of dopamine neurons in cell culture. This will accelerate the development of cell replacement therapies and studies of the biology of dopamine cells.

Lee SH, Lumelsky N, Studer L, Auerbach JM, McKay RD: Efficient generation of midbrain and hindbrain neurons from mouse embryonic stem cells. Nature Biotechnology, 18(6):675-9. 2000.

Hearing Aids. Several Small Business Innovation Research (SBIR) projects have resulted in commercial hearing aid products. The Radiant Beam Array (RBA) was developed as a collaborative effort between Cardinal Sound Labs, Inc, the NIDCD (through SBIR Phase I and II grants), and Starkey Laboratories. The RBA uses six spatially separated microphones with directional signal processing. This technology maximally amplifies sounds directly in front of the listener and minimally amplifies sounds from the sides and rear, thus providing a substantial improvement in the signal to noise ratio available to the user. The result is a greater ability to understand speech in the presence of noise and reverberation. The RBA is worn on the user's chest like a necklace and may be worn under clothing. The RBA is suitable for new and previous hearing instrument users who have a wide range of moderate to profound hearing losses.

Lehr M: Directional Microphone Array for Hearing Aids. Starkey Laboratories. 2000.

Altruism Motivates Participation in Schizophrenia Research. A recent NIH-supported study has helped to characterize previously neglected attitudes of persons with schizophrenia and clinicians toward ethically important aspects of research participation. People with schizophrenia offer highly discerning and complex views, challenging the notion that they may not be competent to provide informed consent for such research. Sixty-three people with schizophrenia and 73 psychiatric faculty and residents were asked to rate 23 attitudes about ethical issues involved in research on schizophrenia. Both groups identified helping others and helping science as important reasons for participation in such research. Schizophrenia patients endorsed the feeling of hope associated with research involvement, a perspective underestimated by the psychiatrists. The psychiatrists also underestimated the acceptance by persons with schizophrenia of physician, investigator, and family influences on decisions about participation. Psychiatrists agreed more strongly than patients that vulnerable populations should be included in research.

Roberts LW, Warner TD, Brody JL: Perspectives of patients with schizophrenia and psychiatrists regarding ethically important aspects of research participation. The American Journal of Psychiatry, 157(1):67-74. 2000.

Assuring Safe Feeding Tube Placement. Every year an estimated one million hospital patients or residents of nursing homes are fed through the use of feeding tubes with stethoscopes or X-rays used to

determine tube placement. Correct placement of a feeding tube is crucial to prevent serious and possibly life threatening complications such as inadvertent administration of tube feeding solutions into the lung. Nurse investigators have discovered that measuring the acid level and bilirubin of aspirated feeding tube contents identifies misplacement of tubes in lungs with 100 percent accuracy. Currently there is no bedside measure of bilirubin. Nurse investigators have extended their previous work and validated a test strip that uses a new colorimetric visual bilirubin scale to assess bilirubin at the bedside. The newly developed scale correctly measured bilirubin values above and below 5 mg/dL in 93.8% of 719 specimens. The combination of acid level and bilirubin continued to identify 100% of the respiratory cases and improved prediction of gastric cases to 98%. Following further testing, refinement and approval by the Food and Drug Administration, assessing the contents of feeding tube aspirates using the visual bilirubin test in combination with acid level has the potential to vastly improve the accuracy of predicting feeding tube location quickly at the bedside and minimizing the chances of serious complications.

Metheny NA, Stewart BJ, Smith L, et al: pH and concentration of bilirubin in feeding tube aspirates as predictors of tube placement. Nursing Research, 48(4):189-97. 1999.

A Database of SNPs Will Shed Light on Human Variation and Disease. A key element necessary to allow medical researchers to make practical use of the human genome sequence is a knowledge of the natural variation in that sequence. Single Nucleotide Polymorphisms (SNPs) constitute the largest component of this variation and play a central role in human disease. NIH has expanded its dbSNP database to include over 800,000 SNPs. The BLAST service for doing sequence similarity searches has also been expanded to include SNP data, opening new avenues of discovery. Scientists can now compare their own newly-acquired sequence data against the entire dbSNP database to quickly locate SNP loci within their gene of interest.

Smigielski EM, Sirotkin K, Ward M, Sherry ST: dbSNP: a database of single nucleotide polymorphisms. Nucleic Acids Research, 28(1):352-5. 2000.

A New Method for Searching for Patterns in Proteins. A major goal of genome sequencing projects is to determine the properties of the proteins encoded by DNA within a genome. These properties determine pathogenicity in some organisms, disease resistance in others and are at the core of many human diseases. When a protein sequence is derived from a gene sequence, many of the properties of the protein can be uncovered directly by finding a similarity to another protein that has already been well-studied. A powerful new tool for performing the search for similarity is IMPALA. IMPALA allows researchers to use their gene-derived protein sequences to search a database of protein patterns. When a match is found, important information as to experimentally- determined biological activity and three-dimensional structure becomes available. By searching a database of

patterns, rather than the usual database of individual protein sequences, similarities of a much more subtle nature can be easily discerned which will allow a more extensive linkage between new sequence data and existing laboratory results.

Schaffer AA, Wolf YI, Ponting CP, Koonin EV, Aravind L, Altschul SF: IMPALA: matching a protein sequence against a collection of PSI-BLAST-constructed position-specific score matrices. Bioinformatics, 15(12):1000-11. 1999.

Complete Genome Data for More Than 700 Organisms Facilitates Comparative Analyses. The Genomes database, a component of the Entrez molecular biology information retrieval service, has expanded its scope to provide access to an unprecedented library of more than 700 complete genome sequences of interest to investigators world-wide. The genomes covered within the Entrez service include those of pathogenic and non-pathogenic viruses and bacteria, plasmids, often implicated in bacterial drug resistance, as well as the yeast and worm genomes used as models in the study of the human genome. Pre-computed bioinformatics analyses developed by NIH include comparisons of many genomes, showing the degree to which they share similar genes. These comparisons make it possible, for instance, to quickly identify a set of genes shared between several types of pathogenic bacteria. Links from sequence to structural information allow researchers to move smoothly from a gene to a structural template to use in modeling studies. The improved methods for organizing and disseminating genomic information on this scale enables the Entrez Genomes service to offer investigations that were previously not possible. For example, using this genomic information can simplify the search for virulence factors in pathogens and facilitate the identification and modeling of drug targets.

Tatusova TA, Karsch-Mizrachi I, Ostell JA: Complete genomes in WWW Entrez: data representation and analysis. Bioinformatics,(7-8):536-43. 1999.

http://www.nlm.nih.gov/research/visible/

Next Generation Internet Implementation to Serve Visible Human Datasets. NIH is supporting a project to develop a Next Generation Internet (NGI) production system to serve Visible Human datasets in novel and educationally useful ways. These include a comprehensive set of interactive 2D and 3D VH browers, featuring arbitrary 2D cutting and 3D visualizations. An interactive WWW navigation engine will be deployed to create and visualize anatomic flythroughs, under haptic control of the user, and also to deliver flythroughs developed by expert anatomists in concert with clinicians. Anatomical labels will enhance these visualization sequences, and enable real time links with appropriate resources on the WWW using XML. As such, this system will complement and extend currently deployed passive WWW information systems with active computational services. This will allow for delivery of several simultaneous high quality digital streams, creating structured medical knowledge using the VH datasets. An experienced evaluation team will measure performance and educational

effectiveness using emerging Advanced Distributed Learning (ADL) principles. Networking experts will provide NGI connectivity and evaluate successes and failures.

http://www.nlm.nih.gov/research/visible/

Remote Treatment Planning System. NIH is supporting the development, implementation, and evaluation of an application to support remote treatment planning for radiation therapy. This application relies on network infrastructure technology for collaboration; on high bandwidth and quality of service (QoS) to support interactive review sessions; and on data privacy and security to protect patient privacy, confidentiality, and data integrity. Review sessions provide a collaborative environment for dosimetrists at the planning site, the oncologists at the care delivery site, and peer reviewers. It utilizes video teleconferencing and a shared view of the images to support treatment planning. The evaluation will measure outcomes at the care delivery site, process improvements at the treatment-planning site, and estimate cost impact on the remote treatment planning process.

http://www.nlm.nih.gov/research/ngiinit.html

Mammography for the Next Generation Internet. NIH is supporting the development of a testbed to demonstrate the feasibility of a national breast imaging archive and network infrastructure to support digital mammography using Next Generation Internet (NGI) technologies. The goal is to improve access and performance of breast cancer screening with an imaging archive that supports storage, retrieval and distribution of breast images for clinical and research purposes and ensures privacy and confidentiality with multilevel security embedded throughout the system. The proposed infrastructure will: 1) support traditional breast screening through the maintenance and distribution of a digital record of prior breast examinations and relevant medical history for primary interpretation and expert consultation; 2) provide the opportunity to maintain and apply computer-aided diagnosis software at central, well-maintained computing resources to studies from all women; 3) provide unique tools for creating educational and training programs; and 4) create an unparalleled opportunity to study and understand many epidemiologic issues in breast cancer through searches of a national breast screening database. NGI technologies will be used to transfer large data files, execute real-time queries, and access information securely. The testbed will demonstrate that quality of service, medical data privacy and security, nomadic computing, network management research and development, and infrastructure technology for collaboration, are NGI technologies that are integral to widespread deployment and optimal utilization of digital mammography.

http://www.nlm.nih.gov/research/ngiinit.html

Multilateral Initiative on Malaria. NIH has led an international effort to provide malaria researchers in Africa with full access to the Internet and the resources of the World Wide Web. This project began with NIH's leadership in the Multilateral Initiative on Malaria in which African scientists identified electronic communication and access to scientific information as critical in the fight against the devastating and economically debilitating effects of malaria in developing countries. Results at the completion of the Alpha Phase and Phase 1: researchers at the Malaria Research and Training Center in Bamako, Mali, are connected by radio waves to their local Internet Service Provider and their colleagues at the CDC/KEMRI and Wellcome/KEMRI sites in Kenya now fully connected via satellite for data, voice, and image. Phase 2 comprises two sites in Ghana which also have full Internet connectivity – Noguchi Institute in Accra and the Navrongo Health Research Center; both sites are engaged in a clinical trial for a malaria vaccine. Phase 2 will be completed in September 2000 when three sites are Tanzania join the network. Partners, in addition to the NIH, included: Centers for Disease Control, Wellcome Trust, World Bank, USAID, Naval Institute of Medical Research, and Walter Reed Army Institute of Research.

Tribal Connections. NIH continues to support a ground-breaking effort to improve the Internet connectivity on selected American Indian reservations and Alaska Native villages. The project began in the Pacific Northwest, where it continues, and is being extended to selected Indian tribes in the Pacific Southwest. By improving connectivity, NIH hopes to facilitate access of Native Americans living in rural, remote areas to health and biomedical information available over the Internet. The NIH project, in collaboration with the Regional Medical Library at the University of Washington in Seattle, and others, is using a community-based infrastructure development approach to help assure that Internet enhancements are responsive to local needs and conditions, and involve the local tribal and village leadership and health community. The project support at each site includes planning and technical assistance, training, and outreach in addition to provision of needed hardware, software, and telecommunications links. NIH is encouraging the involvement of other organizations, such as the Indian Health Service and State telecommunications departments, in order to make best use of scarce resources in a coordinated, sustainable way.

http//:www.tribalconnections.org/

Internet Connectivity Performance Evaluation. NIH continues its research on evaluating the performance of end-to-end Internet pathways involving biomedical institutions and users. The research began with a focus on the performance of the so-called "commodity Internet" and has now shifted to include high-bandwidth Internet connections. The Internet is heavily used to support biomedical research and scientific collaboration, and thus the quality of Internet performance is a major concern. This research has developed a set of methods and metrics for assessing Internet performance, and has applied these tools to evaluate the connectivity of Internet pathways between NIH and selected

research universities, medical libraries, hospitals, and medical researchers in the U.S. and abroad. The research has compared the relative performance of commodity Internet and high-bandwidth links for downloading documents from web sites, searching biomedical databases, and transferring large files of data and images.

Wood FB, Cid VH, Siegel ER: Evaluating internet end-to-end performance: overview of test methodology and results. Journal of the American Medical Information Association, 5(6):528-45. 1998.

New Imaging Instrumentation for Diagnosis of Eye Diseases. Optical coherence tomography (OCT) is a novel non-invasive imaging technique, similar to ultrasound, that promises to have a broad range of applications to the diagnosis and management of ocular disease. The most advanced application is for optic nerve measurements in glaucoma patients. Promising results have been attained with an instrument that has ten micron resolution, and an instrument with five fold greater resolution is now in development. This new OCT is expected to be more reliable and more precise. Increased precision may allow earlier diagnosis as well as more certain identification of progressive changes that impact sight. Reliable assessment of optic nerve damage has been a long sought goal in glaucoma diagnosis and clinical management. And OCT should become an indispensable tool for glaucoma clinicians.

Drexler W, Morgner U, Kärtner FX, et al: In *vivo* ultrahigh-resolution optical coherence tomography. Optics Letters, 24(17):1221-3. 1999.

Intranasal Administration of Naked TGF-B DNA May Be Used to Treat Inflammatory Diseases. Experimenting with mice, investigators demonstrated that nasal spray delivery of DNA from TGF-B, a protein that modulates immune responses, successfully treated experimentally induced colitis (an immune-mediated bowel inflammation) without causing the harmful side effects seen in earlier studies using other modes of administration of this same molecule. However, the reason for this encouraging finding is not yet known and further research is needed to determine whether it is applicable to treatment of human diseases.

Kitani A, Fuss IJ, Nakamura K, Schwartz OM, Usui T, Strober W: Treatment of experimental (trinitrobenzene sulfonic acid) colitis by intranasal administration of transforming growth factor (TGF)-ß1 plasmid: TGF-ß1-mediated suppression of T helper cell type 1 response occurs by interleukin (IL)-10 induction and IL-12 receptor ß2 chain downregulation. The Journal of Experimental Medicine, 192(1):41-52. 2000.

Early Detection of Cancer Cells. More than 85 percent of all cancers originate in the cells that line the internal surfaces of organs throughout the body. These cancers are readily treatable if detected early. Until now, early detection has only been possible after removing a sample of tissue for study, but

researchers at MIT have developed a technique, called laser-scattering spectroscopy, that is able to detect early-stage and pre-cancerous cells in real time without the need for tissue removal. This technique should significantly improve the efficiency of cancer screening.

Backman V, Wallace MB, Perelman LT, et al: Detection of preinvasive cancer cells. Nature, 406(6791):35-6. 2000.

Novel SPECT Ligands to Image the Serotonin Transporter in Human Brain. The serotonin transporter (SERT)is a tiny pump-like component of certain neurons that plays an important role in regulating the amount of the brain's chemical messenger, serotonin, in the synapse. SERT is the site of action of many clinically useful drugs, most notably the selective serotonin reuptake inhibitor (SSRI) class of antidepressants, the prototype of which is fluoxetine, or Prozac. Although an ability to image SERT in humans would aid in understanding how alterations in serotonergic function relate to the pathophysiology and treatment of serious mood disorders, an obstacle has been identifying a ligand – a small molecule that would bind to SERT – that is compatible with SPECT (single photon emission computed tomography) imaging technology. Recently, NIH-funded investigators identified a series of novel small molecules that selectively and strongly bind to the SERT in living brains. Labeled with a radioactive tag, these ligands make it possible to capture SPECT images of serotonin transporter binding sites in living rat and baboon brains. This research provides the groundwork for future studies in humans.

Zhuang ZP, Choi SR, Hou C, Mu M, Kung MP, Acton PD, Kung HF: A novel serotonin transporter ligand: (5-Iodo-2-(2-dimethylaminomethylphenoxy)-benzyl alcohol.. Nuclear Medicine and Biology, 27(2):169-75. 2000.

New Technique Eliminates a Step in Making Genetically-Altered Mice. By transplanting nuclei of genetically-altered mouse embryonic stem cells directly into eggs lacking a nucleus, researchers have eliminated a step needed to produce genetically-engineered animals. Scientists often use such animals to make model systems in which they can study human disease and disorders. Previously, researchers had to introduce a genetically-altered cell into a blastocyst, an early cellular stage in the formation of an embryo, and then breed the resultant animals to produce genetically-pure animals having the desired trait. This breakthrough will significantly shorten the time needed to produce transgenic mice for medical research, as well as provide new insights into the way an egg cell can develop into all of the different cells of the body.

Rideout WM III, Wakayama T, Wutz A, et al: Generation of mice from wild-type and targeted ES cells by nuclear cloning. Nature Genetics, 24(2):109-10. 2000.

Wakayama T, Rodriquez I, Perry ACF, Yanagimachi R, Mombaerts P: Mice cloned from embryonic stem cells. Proceedings of the Natioal Academy of Sciences, 96(26):14984-9. 1999.

RNA Interference: A New Method for Controlling Gene Expression. RNA interference, or RNAi, has become an unexpected, but extremely efficient and versatile means of inhibiting the activity of specific genes. Double-stranded RNA that corresponds to the DNA sequence of the target gene mediates RNAi. Once in the cell, the double-stranded RNA is converted into discrete fragments that bring about cleavage of the corresponding messenger RNA, thereby suppressing expression of the target gene. RNAi has found widespread application as a means of determining gene function in plants and lower organisms whose cells have a true nucleus, and there are indications that it may also operate in vertebrates.

Zamore PD, Tuschl T, Sharp PA, Bartel DP: RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. Cell, 101(1):25-33. 2000.

Grishok A, Tabara H, Mello CC: Genetic requirements for inheritance of RNAi in *C. Elegans*. Science, 287(5462): 2494-7. 2000.

STORIES OF DISCOVERY

MEDLINE and MEDLINEplus: A Continuing Story of Discovery

Medical practitioners and medical researchers both depend on the accumulated wisdom of those who have gone before. It was not so long ago that one could assemble this wisdom only by poring over printed bibliographies, usually the *Index Medicus*, which the NIH began publishing in 1879. Today, virtually all biomedical scientists, many health practitioners, and an increasing number of consumers use a variety of methods to search the MEDLINE database to learn about published research findings. This is the story of how an NIH "discovery" has evolved over the decades to become a service that, in the words of Vice President Al Gore, "may do more to reform and improve the quality of health care in the United States than anything else we have done in a long time."

The pioneering MEDLINE project, begun by the NIH in the early seventies, evolved from the computerized system used to produce the *Index Medicus*, which the NIH had installed in 1964. MEDLINE was the first successful marriage of a large reference database with a national telecommunications network, and it has been called the Model T of online databases: although it would usually get you where you wanted to go, it required a pioneering spirit to master its intricacies and the patience of Job to deal with its idiosyncrasies. Even so, the NIH received more requests than it could handle from medical librarians who wanted to be trained so that they could provide literature search services for health professionals and scientists in hospitals, universities, and laboratories.

The eighties saw the introduction of "Grateful Med," a software program created by the NIH that one could load onto a PC or Macintosh and, equipped with a modem and a password, search MEDLINE right from one's home, office, or laboratory. Grateful Med was eagerly snapped up not only by librarians but by health professionals, scientists, students, lawyers, medical journalists and others, who saw the average charge of $2 per MEDLINE search as a bargain. Trying to find the same information in the printed *Index Medicus* would surely cost much more, in time and effort, if it could be done at all. Another change by then was that most MEDLINE references were accompanied by an abstract.

The nineties, of course, is the era of the Internet and the World Wide Web. The NIH had one of the earliest Federal Web sites (1993) and introduced MEDLINE searching via the "Internet Grateful Med" in 1996. Dr. Michael DeBakey, then a Regent of the Library, and Senator Bill Frist of Tennessee, a surgeon, introduced the new system at a press conference. The following year, Vice President Gore (quoted above) introduced *free* MEDLINE searching via the Web, using a new system called PubMed. Now, for the first time, anyone with access to the Web could search through an immense database of references and abstracts to 11 million medical journal articles. In fact, PubMed and Internet Grateful Med had both simplified MEDLINE searching to the point where the public encountered no difficulty at all in retrieving relevant references on any biomedical subject from the literature. Since about 30

percent of all MEDLINE searches (now totaling about 250 million a year) are being done by consumers, this presented the NIH with a wonderful opportunity. Why not create a service that not only will provide selective MEDLINE results that are useful to the consumer, but also link the Web user to the authoritative, full-text health information being put out by NIH Institutes and by a variety of non-Federal sources? Such a service, called MEDLINE*plus*, was introduced in October 1998.

Where did such "authoritative" information come from? Since the NIH publishes a wealth of consumer health information based on the medical research it sponsors, it was natural to start there. Also "recruited" were professional medical societies and voluntary health agencies, many of which issue, without commercial or business motive, authoritative information that the public can trust. With help from members of the National Network of Libraries of Medicine across the country, NLM information specialists have selected and organized this information and extensively cross-linked it. Because the name "MEDLINE" had a quarter-century exposure to the health professions, and because that database was now also increasingly known to the public, the NIH called the new service MEDLINEplus (http://medlineplus.gov).

In the year and a half since its introduction, the service has grown tremendously, both in terms of its coverage of health and its usage by the public. As of June 2000, MEDLINEplus was being consulted some 2 million times each month. The original two dozen "health topics," containing detailed consumer information on various diseases and health conditions, have increased to 400. Other information available through MEDLINEplus: links to more than 5,000 current clinical trials, medical dictionaries and an encyclopedia, detailed information about prescription drugs, information in Spanish, directories of health professionals and hospitals, links to organizations and libraries that provide health information for the public, and preformulated searches of MEDLINE on various aspects of the health topics to find recent articles (in journals likely to be available) that will be of interest to consumers seeking information for self or family.

The NIH has also learned that health professionals of all kinds are finding MEDLINEplus to be an excellent source of information. Many physicians use it to keep up-to-date on medical subjects outside of their specialty. Others are referring their patients to MEDLINEplus for up-to-date and authoritative information about their health conditions.

In 2000, NIH made "outreach awards" to more than 50 medical libraries and other institutions around the Nation to enable them to work with local, state, and regional organizations to promote public access to reliable health information. In this effort, medical librarians from the National Network of Libraries of Medicine are volunteering to work with the public librarians and teach them about MEDLINEplus and other electronic sources of information.

To sum up, in just the last few years there has been a remarkable change in how consumers seek health information for their personal use. The NIH is riding the crest of this wave and is adapting its extremely

popular electronic database, MEDLINE, to serve the public as well as it has served scientists and health professionals for three decades.

Genome-scale Analyses: cDNA and Tissue Chips

The newfound abundance of genomic information is propelling scientists out of the pattern of studying genes individually; now, scientists are able to monitor thousands of genes at a time. For such large-scale analyses, miniaturized "chip" technologies, also called microarrays, can be rapid, efficient, and economical. All microarrays share this characteristic: they permit researchers to examine many elements in parallel. NIH has supported the development of the major microarray technologies in use, including chips sold commercially, and has promoted communication among researchers in the field.

The flood of data emerging about DNA and genes has required new ways of sorting through the information to find the telling details that will illuminate how the cells that make up living organisms function. Microarrays are being used for many different applications. Some microarrays gauge how active different genes are in different kinds of cells; others let researchers track the molecular changes in tumor cells as cancer progresses. Capturing holistic views of changes within cells has begun to elucidate the signalling pathways that are altered and distinguish a cancerous cell from a non-cancerous cell. Such insights may provide the identification of early, presymptomatic changes in cells and thus, rational therapeutic targets for the treatment of cancer.

Expression arrays, which chart gene activity, have been among the most productive chips so far. To make an expression array, robots spot fragments from thousands of genes onto a region of a single glass microscope slide about the size of a postage stamp. This genomic approach to understanding the impact of an altered genetic code within a cell is analogous to listening to a full orchestra playing a symphony, rather than an individual instrument. While each cell contains the score for each of the approximately 50,000-100,000 or so available instruments (the protein products of genes), it only utilizes a fraction of those. Each different cell type (e.g., brain or muscle cell) will utilize a different compliment of 10,000 or so instruments. While much can be learned from studying an individual instrument, significant limitations exist. Our knowledge grows dramatically when we begin to examine the networks of pathways that are affected by a single alteration in the genome.

An international team led by scientists at NIH recently used this approach in discovering a genetic "signature" that may help explain how malignant melanoma can spread to other parts of the body. It was one of the first large-scale studies on cancer genetics to be made possible by the wealth of information generated by the international Human Genome Project. The research helps explain the mechanisms underlying the metastasis of melanoma. And it may ultimately point the way to better diagnosis and treatment of this increasingly common and often fatal disease.

Using gene expression profiling, the researchers were able to find a genetic signature, or set of differences in genes, that for the first time divided patients with advanced melanoma into subgroups. Almost half a million measurements were taken on nearly 7,000 different genes in melanoma tumors from 40 patients. Computers running sophisticated statistical software were then used to analyze the

data from the chips in order to find hidden patterns in gene expression among the tumor samples. Nineteen cancers were found to be very similar in gene expression, differing from the rest of the tumors in the expression of roughly 500 genes. According to the patient histories, tumors in this cluster tended to be less aggressive, suggesting they metastasized less rapidly. Such classification of cancer on a molecular level offers the possibility of more accurately determining the prognosis of a particular patient's tumor, based on his or her genetic makeup. It also offers the hope of tailoring therapies to the individual.

In order to determine the importance of any gene in a more physiological setting, a second kind of array, called the tissue microarray, can confirm the importance of each gene that emerges as a candidate. NIH researchers have developed a way of arranging some 1000 tiny cylindrical tissue biopsies in a small paraffin block. Thin slices cut from this block can be mixed with a probe that binds to a specific gene or gene product to allow researchers to visualize gene number, activity or subcellular localization of proteins in hundreds of different tissues simultaneously. Tissue arrays permit researchers to examine the molecular details of many different healthy tissue types or in different stages of disease. This past year, NIH researchers combined cDNA and tissue microarray technologies to identify molecular alterations associated with the progression of human breast cancer.

First, researchers applied their new double-chip approach to human breast cancer cells grown in culture to identify genes that were over expressed. Then, tissue chips were constructed, containing tiny dot-sized samples from over 600 breast cancer tumors, and used to determine how frequently these genes were over-expressed in cancers from actual patients.

One of the amplified genes, HER-2, on human chromosome 17, was identified 10 years ago as one of the genes involved in the initiation and progression of breast cancer. As expected, patients whose tumor samples had amplified HER-2 had a poor survival. However, the researchers were surprised to discover on chromosome 17 another gene that was also over-expressed. Because this gene, previously identified by other researchers and called S6K, produces a protein that carries out several important biological functions in the proliferation of cells, it represents a good candidate for a cancer-related gene. Breast cancer tissue with amplified S6K were also associated with a poor prognosis. For those women with breast cancer whose HER-2 and S6K genes were both over-expressed, survival was even worse. As a follow-up to this study, the researchers are now using the DNA chip and tissue chip techniques to look for other genes near S6K on chromosome 17 that may also be amplified in breast cancer tumors.

In all their various forms, microarray technologies can support the study of genetic complexity and are becoming increasingly common as a “genome perspective,” one that considers the entire DNA code of an organism, takes root in biomedicine.

Single Nucleotide Polymorphisms (SNPs): New Tools for Tracing Inherited Diseases

A key aspect of research in genetics is associating sequence variations with heritable diseases. The most common variations are single nucleotide polymorphisms (SNPs), which occur approximately once every 100 to 300 bases. Because SNPs are expected to facilitate large-scale genetic association studies, there has recently been great interest in SNP discovery and detection. The identification of SNPs has accelerated dramatically in the past year, due in large part to the availability of working draft sequence of the human genome.

In FY1999, NIH organized the establishment of the DNA Polymorphism Discovery Resource (PDR) to allow researchers to look for SNPs in a common set of samples. The PDR is now the major resource being used to look for SNPs. It consists of 450 DNA samples collected under strict ethical guidelines from anonymous unrelated United States residents of diverse ethnic backgrounds(http://www.nhgri.nih.gov:80/AboutNHGRI/Der/variat.htm).The NIH has funded studies to allow researchers to look for SNPs in a common set of samples. This permits the accumulation of a data set that can then be used in association studies to identify inherited disease risks.

This effort is complemented by the SNP Consortium (TSC), a non-profit entity whose mission is to develop a high-quality SNP map of the human genome and to make the information related to these SNPs available to the public without intellectual property restrictions (http://snp.cshl.org/). The project started in April 1999 and is anticipated to continue until the end of 2001. The SNP Consortium's members include the medical research charity The Wellcome Trust; 10 pharmaceutical companies including AstraZeneca PLC, Aventis Pharma, Bayer AG, Bristol-Myers Squibb Company, F. Hoffman-La Roche, Glaxo Wellcome PLC, Novartis, Pfizer Inc, Searle (now part of Pharmacia), and SmithKline Beecham PLC; Motorola, Inc.; IBM, and Amersham Pharmacia Biotech. Academic centers including the Whitehead Institute for Biomedical Research, Washington University School of Medicine in St. Louis, the Wellcome Trust's Sanger Centre, Stanford Human Genome Center, and Cold Spring Harbor Laboratory, are involved in SNP identification and analysis.

In July 2000, the HGP and TSC announced a collaboration to accelerate the construction of a higher-density SNP map and enhance the utility of human working draft sequence. At the same time, the data generated will help improve the "working draft" itself. Three academic genome research centers – the Whitehead Institute for Biomedical Research, Washington University School of Medicine, and the Sanger Centre – will participate in this collaboration.

Researchers have recently developed a simple but powerful method, called reduced representation shotgun (RRS) sequencing, for creating SNP maps. RRS facilitates the rapid, inexpensive construction of SNP maps. This technological advance in combination with a dense SNP may should markedly enhance attempts to unravel the genetic contribution to common disease.

The centers will isolate two million DNA fragments (each about 6,000 base pairs long) from the human genome and determine the sequence of approximately 500 base pairs at both ends of the fragments, resulting in paired sequences of known distance from each other. The sequences then will be compared to human genome DNA sequences already in GenBank to identify SNPs. In addition, the paired-end information will help span some gaps in the human genome "working draft," enhancing the value of the draft. This paired-end approach has been used to advantage in the sequencing of the genomes of lower organisms, such as *E.coli* and *Drosophila melanogaster*.

The DNA to be sequenced will come from 24 anonymous, unrelated donors with diverse geographic origins from the PDR, making the new sequences a rich source of SNPs. As SNPs are identified, they will be validated, mapped, and deposited in the publicly accessible NIH database, dbSNP.

The SNP Consortium will file provisional patent applications on newly identified and mapped SNPs solely to establish the dates of discovery, but no patents will be allowed to issue, keeping the data freely available for the unrestricted use of researchers worldwide.

Through this collaboration, the SNP Consortium will be able to contribute about three times as many SNPs to the public domain than otherwise would have been possible under TSC's original scientific plan which had been to identify 300,000 SNPs and map at least 150,000 of these SNPs, evenly distributed throughout the genome. An exponential increase in the amount of human genetic sequence data that has recently become available from the Human Genome Project has enabled the consortium to proceed at a much faster pace than originally envisioned. To date, the consortium has identified over 140,000 SNPs and mapped 102,719 SNPs. With the Human Genome Project collaboration, the total number of validated and useful SNPs mapped may exceed 1,000,000 by December 2000.